

## THESIS / THÈSE

### MASTER EN SCIENCES DE GESTION À FINALITÉ SPÉCIALISÉE EN BUSINESS ANALYSIS & INTEGRATION

L'efficience informationnelle du marché des paris sportifs

Analyse de la prévision des cotes et de l'impact des news

Leduc, Jordan

*Award date:*  
2019

*Awarding institution:*  
Université de Namur

[Link to publication](#)

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



**L'efficience informationnelle du marché  
des paris sportifs : Analyse de la prévision  
des cotes et de l'impact des news**

**Jordan LEDUC**

**Directeur: Prof. P.GIOT**

Mémoire présenté  
en vue de l'obtention du titre de  
Master 120 en Sciences de gestion,  
à finalité spécialisée

**ANNEE ACADEMIQUE 2018-2019**

## **Avant-propos**

Etant passionné par la finance mais également par le monde du sport, ce mémoire représente une opportunité unique de pouvoir mettre en application mes acquis théoriques dans un sujet de recherche qui m'intéresse et que j'affectionne. Toutefois, il n'aurait pas pu voir le jour sans l'aide précieuse de plusieurs personnes que je tiens, ici, à remercier.

Tout d'abord, je souhaiterais souligner la contribution du groupe Ardent et plus précisément de Monsieur Etienne Brunson ainsi que de Monsieur Julien Becquart qui m'ont permis de comprendre les mécanismes qui émanent des paris sportifs tout en me procurant des données exploitables dans le cadre de ce mémoire. Leur support tout au long de la rédaction de mon mémoire m'a, par ailleurs, été d'une aide précieuse.

Mes remerciements s'adressent ensuite à mon promoteur, Monsieur Pierre Giot, qui a su m'orienter de manière efficace dans mes recherches et qui a fait preuve de disponibilité pour répondre à l'ensemble de mes questions.

Je souhaite également exprimer ma gratitude auprès de Marie-Gabrielle Boxus, Philippe Chevalier, Frédéric Vrins, Alexandre Prospero, Dinos Stranomitis, Alan Weber et Renato Bazzarini avec qui j'ai eu la chance d'échanger, et qui ont tous, d'une manière ou d'une autre, contribué à l'élaboration de ce mémoire.

Enfin, j'aimerais remercier ma famille et mes amis qui ont toujours su me motiver, me conseiller et m'encourager durant la réalisation de mon mémoire.

## Table des matières

Avant-propos .....	2
Table des matières .....	3
Introduction .....	5
Première partie : Description théorique des concepts clés .....	8
Chapitre 1 - Le marché des paris sportifs : Fonctionnement et mécanismes.....	8
Introduction aux cotes.....	8
Mécanismes.....	8
Détermination de la cotation.....	12
Les formes alternatives de paris sportifs.....	15
Chapitre 2 - L'efficacité informationnelle du marché des paris sportifs .....	17
Introduction à l'efficacité informationnelle .....	17
La forme faible de l'efficacité.....	18
La forme semi-forte de l'efficacité .....	18
La forme forte de l'efficacité .....	19
Revue littéraire sur l'efficacité du marché des paris sportifs.....	20
Deuxième partie : Tests empiriques et analyses de l'efficacité.....	22
Chapitre 3 - Méthodologie.....	22
Analyse graphique .....	26
Runs Test .....	27
Régression linéaire.....	29
Chapitre 4 – Test de l'efficacité faible pour le basketball .....	31
Analyse du cas n°1 : Mise sur la victoire des Miami Heat .....	31
Analyse du cas n°2 : Mise sur la victoire des New Orleans Pelicans .....	36
Analyse du cas n°3: Mise sur la défaite des Chicago Bulls .....	38
Analyse du cas n°4: Mise sur la défaite des Houston Rockets .....	41
Chapitre 5 – Test de l'efficacité faible pour le baseball .....	44
Analyse du cas n°1 : Mise sur la victoire des Houston Astras.....	44
Analyse du cas n°2 : Mise sur la victoire des Philadelphia Philies .....	47
Analyse du cas n°3 : Mise sur la défaite des Boston Red Sox.....	50
Analyse du cas n°4 : Mise sur la défaite des Colorado Rockies .....	52
Chapitre 6 - Influence des news sur le marché des paris sportifs.....	56

Influence des informations publiques sur la cote « live ».....	58
Vitesse d'ajustement de la cote en live .....	62
Conclusion.....	65
Conclusion sur l'effcience faible.....	65
Conclusion sur l'effcience semi-forte .....	66
Conclusion générale et perspectives .....	66
Bibliographie .....	68
Annexes .....	73

## Introduction

Alors que la portée des paris ne cesse de prendre de l'ampleur, la variété et l'originalité de ceux-ci ne cessent d'étonner. La nouvelle n'aura échappé à personne tant elle a été médiatisée, il est désormais possible de parier sur la couleur du chapeau d'une telle personnalité ou sur le nom d'un futur enfant princier. L'intérêt pour nous de nous focaliser sur les paris sportifs provient du fait que, sur ce marché, les événements ont généralement une durée très limitée par rapport à d'autres marchés prédictifs. En effet, alors que les paris associés au nom du prochain premier ministre français ou à l'année du premier pas de l'homme sur Mars peuvent prendre quelques années, un pari concernant un événement sportif ne dure généralement que quelques heures, ou tout au plus que quelques jours.

Par conséquent, le marché des paris sportifs est connu pour connaître une volatilité importante de ses prix dû à la brièveté de ses événements. En outre, la popularité grandissante de ce marché découle sur un volume impressionnant, à tel point que le volume des mises sur certains événements sportifs est parfois supérieur au volume d'un actif boursier prisé. Ces deux facteurs constituent de ce fait, un cadre de recherche idéal. Par ailleurs, au regard des similarités qu'il possède avec les marchés financiers, ce marché s'avère être un terrain d'étude idéal à l'efficacité informationnelle. Selon VAUGHAN WILLIAMS (1999), il pourrait même mieux convenir que les marchés financiers dans l'examen des différentes hypothèses d'efficacité car, contrairement à la plupart des marchés financiers, les paris ont, comme nous venons de le préciser, une durée limitée dans le temps au terme de laquelle leur valeur devient observable mais également parce que les parieurs connaissent généralement leur potentiels gains futurs. A noter que le nombre important d'informations publiques circulant dans ce marché est aussi un atout indispensable quant à l'étude de l'efficacité de ce marché.

Apportons quand même une nuance à tout ceci. Bien que les deux marchés soient relativement similaires, la raison pour laquelle ces deux marchés existent est tout à fait différente. Nous ne reviendrons pas sur une distinction détaillée de l'origine de ces deux marchés, mais le marché des paris sportifs, lui, est beaucoup plus commercial que financier. En effet, l'approche du marché des paris sportifs est un peu la même que pour le Lotto ou pour le poker ; le seul avantage potentiel (en plus du plaisir de jouer lui-même) est son enrichissement personnel et l'action de miser ou non n'affecte pas (ou rarement) des personnes tierces. A contrario, l'investissement dans un actif financier a lui une implication

beaucoup plus importante. Dans les deux cas, cependant, il s'agit effectivement de prédire l'évolution future, soit d'un cours de bourse, soit d'un évènement sportif.

L'efficience informationnelle du marché des paris sportifs est d'autant plus intéressante à analyser, qu'il a connu récemment une évolution drastique, et ce, grâce à internet. Effectivement, alors que la naissance des paris sportifs ne date pas d'hier, l'émergence et le développement de ceux-ci sont bien plus récents que l'on peut l'imaginer. A l'image de l'évolution frappante qu'a connu le monde de la Bourse avec l'automatisation de la cotation<sup>1</sup> et l'instauration de la cotation boursière en continu qui en découle, les paris sportifs ont également changé de manière significative au cours des dernières années. « Une logique de cotation dynamique comparable à celle des marchés financiers s'installe alors progressivement chez le bookmaker » MACKENZIE et MILLO (2003)

Comme le précise TRESPEUCH (2011) « internet introduit de l'instantanéité dans le processus » et, de ce fait, a permis au marché des paris sportif d'être plus efficient qu'il ne l'était auparavant. En effet, auparavant, à la veille de l'automatisation et d'internet, les cotations ne variaient pas de manière continue : elles étaient cotées à un moment de la journée selon le procédé du « fixing ». Cela signifie que chaque évènement sportif ne se voyait attribué que deux ou trois cotes maximum par jour.<sup>2</sup> Puisque le réseau était auparavant essentiellement « physique » (le seul moyen de parier était de se rendre dans un point de vente), la modification de la cotation représentait un réel parcours du combattant comparé à ce qui se fait aujourd'hui ; le courtier devait imprimer les listes actualisées et les transmettre aux gérants des différents points de vente pour que ces derniers puissent afficher les nouvelles cotations. Inutile de préciser que l'efficience du marché des paris sportifs était à l'époque toute relative par rapport à ce qu'elle est devenue. Tester les hypothèses d'efficience de ce marché s'avère ainsi être beaucoup plus pertinent à l'heure actuelle, de par les similitudes observées avec l'évolution du marché financier, mais également de par le nombre de données et d'informations qu'il est possible de collecter de nos jours.

---

<sup>1</sup> Notamment avec l'instauration du système CATs (Computer Assisted Trading system) en 1986

<sup>2</sup> De nos jours, le bookmaker peut à tout moment changer la cotation via son interface en ligne, et ce, sans aucune limite.

C'est donc dans ce contexte d'évolution constante, de proximité avec le marché financier et d'efficience que s'inscrit ce mémoire. Organisé en 6 chapitres différents, ce travail propose, en effet, une analyse sur l'efficience informationnelle du marché des paris sportifs. Les particularités de celui-ci résident dans l'utilisation de données récentes et suffisamment étendues que pour être pertinentes, la diversité des tests effectués (le « run test » par exemple) ainsi que des analyses menées sur trois sports distincts.

Tout d'abord, la question du fonctionnement et des mécanismes des paris sportifs apparaît comme centrale dans la compréhension des tests entrepris dans le cadre de mes recherches et sera dès lors évoquée au cours du premier chapitre.

Ensuite, le second chapitre aura pour but de présenter l'efficience informationnelle et ses différentes formes, tout en exposant la revue littéraire qui en dépend.

La troisième partie mettra l'accent sur le développement et la méthodologie appliquée dans le cadre de la réalisation des tests empiriques. Ce chapitre permettra de faire comprendre au lecteur les différentes démarches qui ont été entreprises pour mener à bien les analyses.

Le quatrième ainsi que le cinquième chapitre seront dédiés à l'analyse de l'efficience faible du marché des paris sportifs. Divers tests seront effectués dans le but de vérifier s'il est possible pour un parieur de se baser sur les rendements passés pour prédire les rendements futurs.

La sixième partie s'interrogera, elle, sur l'efficience semi-forte et sur la capacité des bookmakers à ajuster leur cotation rapidement en fonction des informations publiques qui sont émises. On tentera alors de démontrer que les bookmakers proposent des cotations qui intègrent de manière instantanée les informations publiques.

Enfin, nous apporterons des conclusions émanant des observations faites au cours des trois derniers chapitres, tout en proposant des perspectives de recherches futures.



# **Première partie : Description théorique des concepts clés**

## **Chapitre 1 - Le marché des paris sportifs : Fonctionnement et mécanismes**

### **Introduction aux cotes**

Les côtes ont une importance considérable dans les paris sportifs du fait qu'elles déterminent aussi bien le gain potentiel des parieurs que ceux des bookmakers. En effet, « une cote est associée à chacun des résultats possibles d'un pari. Les cotes informent les parieurs sur le gain potentiel qu'ils peuvent récupérer en cas de réalisation de chacun de ces résultats (PAVEE, 2010). Cette note représente en fait le montant que ce bookmaker est prêt à vous donner dans le cas où la mise est bonne. C'est pourquoi il est important de détailler la logique inhérente aux cotes avant d'entamer la partie analytique.

Il existe trois manières d'exprimer les cotes d'un pari - les cotes fractionnelles, les cotes décimales et les cotes "MoneyLine". De manière générale, on dira que les cotes fractionnelles sont utilisées au Royaume-Uni, les cotes décimales ont pour habitude d'être présentes dans les marchés européens, canadiens ou australiens alors qu'on a davantage recours à des côtes dites "MoneyLine" aux Etats-Unis. Attention cependant qu'il s'agit plus d'une règle tacite que d'une réglementation à proprement parler. Si un bookmaker anglais veut imposer une cote fractionnelle sur le marché européen, il en a tout fait le droit, mais il ne le fera normalement pas car le public ciblé n'en a pas l'habitude et aura tendance à se méfier d'une cotation qu'il ne maîtrise pas.

Au cours du point « Mécanismes », ci-dessous, nous allons omettre volontairement l'explication du modèle fractionnel et celui du Moneyline pour nous concentrer exclusivement sur la côte décimale puisque c'est celle-là qui sera utilisée dans le cadre de ce mémoire.

### **Mécanismes**

Si l'on admet qu'un pari quelconque pourrait être défini comme une action visant à prendre un risque sur un événement incertain, le pari sportif consisterait donc à miser une somme d'argent (risque financier) sur la réalisation d'un événement sportif spécifique. Au fil des années, les paris sportifs se sont développés d'une telle manière qu'il est maintenant à peu près possible de parier sur presque chaque fait de match (nombre de buts, de cartons rouges ou de penaltys en football, nombre de secondes entre le premier et le deuxième en cyclisme,

nombre d'aces ou de points gagnés au tennis, etc...) et non plus seulement sur l'issue du match.

Néanmoins, dans la plupart des sports collectifs, le pari le plus classique demeure la mise sur le résultat final du match. Dans un tel scénario, il existe trois possibilités de résultats et donc trois possibilités de mises :

Victoire | Nul | Défaite

Que l'on peut aussi voir comme étant :

Victoire Equipe1 | Nul | Victoire Equipe2

Comme nous venons de le signaler, le pari implique la prise de risque et, tout comme en matière de spéculation boursière, le gain potentiel est proportionnel à cette prise de risque.

En terme de paris sportifs, c'est la cote qui représente ce rapport au risque. En théorie, ces cotes ( $C_i$ ), proposées par les bookmakers, sont l'inverse des probabilités ( $P$ ) de réalisation des résultats auxquelles elles sont associées (Toulemonde, 2011)

$$C_i = 1/P \quad (1.1)$$

L'interprétation découlant de cette formule s'avère être relativement aisée. Au plus la probabilité d'occurrence de cet événement est importante, au plus la cote et donc le gain potentiel qui y est associé sera faible, et inversement, plus la probabilité d'occurrence de l'événement est faible, au plus la cote et le gain potentiel qu'elle représente seront élevés.

En d'autres termes, plus la cote sous-jacente à un certain événement tend vers 1, plus l'événement est considéré comme susceptible de se produire.

Ce qu'il est intéressant de remarquer est que si l'on effectue un test sur un pari existant on se rend compte que le pourcentage est souvent supérieur à 100%. En effet, la somme des probabilités de marché sur tous les différents scénarios possibles d'un événement sont généralement supérieures à 100% en raison de la marge bénéficiaire prise par le bookmaker, qu'on appelle plus communément « over-round ».

Prenons un exemple actuel en considérant l'annexe A-1 qui représente la cote de pré match du site Betfirst lors de la confrontation de Barcelone contre le Real Madrid le 6 février 2019.

Dans cette situation, la cote de Barcelone est de 1.74, le match nul est à 4.05 et elle est de 4.50 pour le Real Madrid. En appliquant la formule (1.1) à l'envers, on arrive respectivement à 57,47%, 24,59% et 22,73% de chance de réussite pour chaque possibilité de mise. En additionnant ces probabilités, on arrive à un total de 104.79% (et non de 100% comme on pourrait l'imaginer).

A partir de ce constat, il est possible de calculer un « taux de retour aux joueurs » (TRJ), également appelé « taux de déboursement », ou « payback » représentant la « part des mises qui sera réservée aux parieurs gagnants » (PAVEE, 2010), en appliquant la formule suivante :

$$TRJ = 1 / \sum P_i \quad (1.2)$$

Pour les données reprises dans l'annexe A-1, on obtient un taux de restitution de l'ordre de 95,43%. Cela signifie que les bookmakers « biaisent les cotes à la baisse » (TOULEMONDE, 2011). Dans ce cas précis, ce sont près de 5% qui sont prélevés par les bookmakers pour financer leurs propres frais de fonctionnement, pour payer les redevances à l'Etat ou pour alimenter les bénéfices leur propre société.

La question que l'on se pose est de savoir comment l'on pourrait procéder pour calculer les vraies probabilités (ProbR) d'occurrence des événements, puisque celles-ci sont généralement biaisées par la taxe prélevée par le bookmaker.

Pour ce faire, il suffit d'appliquer la formule suivante :

$$ProbR = (1/C_i) * TRJ \quad (1.3)$$

Dans le cas présent on aura comme réelle probabilité d'occurrence (c'est-à-dire en enlevant la marge prise par le bookmaker) :

$$\text{Barcelone : } (1/1.74) * 95,43 = 55.85\%$$

$$\text{Nul : } (1/4.05) * 95,43 = 23.56\%$$

$$\text{Real de Madrid : } (1/4.4) * 95,43 = 21.69\%$$

Ce qui, au total, fait bien 100%.

Les cotations qui devraient normalement s'appliquer si le bookmaker décidait de ne pas instaurer de marge bénéficiaire seraient donc de l'ordre de :

$$C_{\text{Barcelone}} = 1/0.5585 = 1.79$$

$$C_{\text{nul}} = 1/0.2356 = 4.24$$

$$C_{\text{Real de Madrid}} = 1/0.2169 = 4.61$$

Attention que ceci n'est vrai que si le bookmaker décide de répartir de manière équitable ses bénéfices entre le favori et l'outsider. En effet, cela peut paraître un peu étrange, mais lorsqu'un joueur ou une équipe fait clairement office de favori, et l'adversaire d'outsider, la marge est répartie de manière à surévaluer l'outsider et à sous-évaluer le favori. Ce biais est bien connu, et largement documenté, dans le monde des paris sportifs sous le nom de « favourite-longshot bias ».

Ainsi, alors qu'on a vu que les cotes étaient biaisées à la baisse suite à l'instauration d'une marge bénéficiaire de la part du bookmaker, il est prouvé que la marge pèse davantage sur l'outsider que sur le favori. Dans notre cas, Barcelone étant, au vue de la cote proposée, favori du match, sa cote ne sera pas trop impactée. Par contre, le Real de Madrid étant outsider, on peut supposer que sa cote a largement été amputée par la marge bénéficiaire du bookmaker.

Le tableau ci-dessous propose un récapitulatif de ce qui vient d'être dit. Les cotes observées chez le bookmaker Betfirst sont de respectivement 1.74, 4.05 et 4.5. Sans la marge bénéficiaire prélevée par le bookmaker, les cotes seraient de respectivement 1.79, 4.24 et 4.61 si et seulement si la marge est uniformément répartie pour les 3 scénarios (33.33% de marge déduite à chaque scénario). Par contre, au vue de la littérature et du favourite-longshot biais, on sait que la cote du joueur ou de l'équipe outsider a été davantage réduite que celle du favori. Une estimation vraisemblable mais approximative des cotes réelles, sans marge bénéficiaire, est donc proposée en troisième ligne du tableau 1.1.

**Tableau 1.1 : Aperçu des cotes observées avec la marge bénéficiaire et des cotes estimées sans la marge bénéficiaire**

	Barcelone	Match Nul	Real de Madrid
Cotes observées	1,74	4,05	4,5
Marge équitable	1,79	4,24	4,61
Estimation Favourite-longshot biais	1,76	4,12	5

Nous n'étudierons ici pas en détail ce phénomène qui a déjà été prouvé à de maintes reprises et nous concentrerons davantage sur la prévisibilité des rendements futurs lors de notre analyse de l'efficience faible.

## Détermination de la cotation

Sur base de ce que nous avons dit ci-dessus, on pourrait supposer qu'il est aisé de fixer les cotes des paris sportifs. En effet, le bookmaker pourrait se contenter d'estimer la probabilité d'occurrence de chacun des résultats d'un pari sur base de ces intuitions, et appliquer la formule (1.1), tout en n'oubliant pas d'y ajouter sa marge bénéficiaire à hauteur de 5-10%. Néanmoins, dans la pratique, la détermination des cotes s'avère être beaucoup plus complexe.

Il est possible de diviser le processus de détermination des cotes en deux phases distinctes : Le calcul d'une cote approximative puis l'ajustement de la cote.

### Phase n°1 : Mécanisme de détermination d'une cote initiale théorique.

Dans un premier temps, l'objectif est d'estimer les probabilités mentionnées plus haut. Pour y parvenir, les bookmakers utilisent différents algorithmes très évolués et complexes, que chaque bookmaker préfère garder secret, permettant, sur base de données historiques, d'estimer les chances de chaque événement de se produire. En effet, en consultant leurs propres bases de données, les bookmakers ont accès à l'entièreté des données statistiques qui ont eu lieu dans le passé. De cette manière, les bookmakers vont regarder si l'événement sportif s'est déjà déroulé par le passé et si tel est le cas, ils consulteront les cotes qui avaient été fixées à l'époque ainsi que le résultat final de l'événement en question. Les premières études réalisées sur ce sujet (mentionnons par exemple MORONEY en 1956) montrent que la modélisation statistique permet d'obtenir de bonnes estimations concernant les probabilités d'occurrence. Ces bases de données constituent donc un réel must pour tout bookmaker qui se respecte car elles permettent de proposer aux bookmakers les cotes qu'il est le plus judicieux d'adopter (aussi bien en terme d'attractivité pour le parieur qu'en terme de rentabilité pour le bookmaker), et ce, pour n'importe quel événement de n'importe quel sport. Néanmoins, les bookmakers tiennent quand même compte de la validité temporelle des résultats ; Un événement similaire s'étant déroulé il y a une trentaine d'années aura forcément moins de poids qu'un événement similaire récent.

### Phase n°2 : Le mécanisme d'ajustement des cotes

Alors que la première phase est donc entièrement automatisée (le bookmaker ne fait que de constater les probabilités qui émanent des analyses faites statistiquement par la machine), la deuxième nécessite quant à elle une intervention humaine. En effet, les bookmakers «

consultent leurs serveurs uniquement pour avoir une base de calcul pour les côtes » (TOULEMONDE, 2011) et on verra par la suite que l'importance de la première phase est toute relative par rapport à la deuxième qui apparaît comme étant déterminante.

Pour le bookmaker, le processus d'ajustement des côtes est réalisable manuellement et de manière instantanée puisqu'il lui est possible de modifier la cotation via son logiciel de cotation.

Cet ajustement manuel des cotes dépendra principalement de trois facteurs distincts.

Le premier facteur nécessitant une intervention humaine relève du fait qu'un comparatif doit absolument être fait entre les cotes du bookmaker et de ses concurrents. On parle alors de la réalisation d'un « consensus » (TOULEMONDE, 2011) avec les autres bookmakers. L'enjeu de cette comparaison est double. Alors qu'elle permet en premier lieu de vérifier que la cote initiale théorique résultant de leurs bases de données est partagée par les autres bookmakers, elle accorde aussi au bookmaker l'assurance d'avoir une cote qui est compétitive. Le rôle du bookmaker représente donc, comme nous l'avons signalé auparavant, un véritable travail d'équilibriste consistant à trouver la cotation optimale, c'est-à-dire celle qui est assez basse pour pouvoir en retirer un maximum de bénéfice mais assez haute que pour pouvoir attirer le parieur potentiel en se différenciant des autres bookmakers. Par ailleurs, ce « consensus » tenu par les bookmakers leur permet d'éviter que les parieurs établissent des stratégies d'arbitrages qui pourraient mettre à mal l'efficience du marché des paris sportifs.<sup>3</sup>

Ensuite, puisque l'intérêt principal du bookmaker est de réaliser un profit, il va faire en sorte d'ajuster la cotation pour que le risque encouru soit minimum. Autrement dit, il va avoir pour objectif d'égaliser le volume des paris placés sur l'ensemble des scénarios d'un événement sportif. Certains auteurs bien connus de la revue littéraire des paris sportifs comme VERGIN et SOSIK (1999), MALIOS (2000), CANTINOTTI et al. (2004) ou encore GODDARD et ASIMAKOPOULOS (2004) stipulent en effet que « l'objectif des bookmakers est avant tout de récupérer des mises équilibrées sur les différents résultats possibles d'un pari donné ». De la sorte, les bookmakers ne perdent jamais. Peu importe le résultat, si les mises sont égales sur tous les scénarios d'un événement, le risque est réparti et le bénéfice est donc assuré. Ainsi, «

---

<sup>3</sup> Si un bookmaker vient à opter pour une stratégie de différenciation relative aux cotations, le marché des paris sportifs tout entier peut être mis à mal puisqu'un parieur pourrait miser sur l'ensemble des scénarios possibles d'un événement et réaliser un profit. En effet, si les bookmakers ne s'alignaient pas un minimum sur la cote des bookmakers concurrents, un parieur pourrait par exemple très bien parier simultanément sur la victoire, le nul et la défaite d'une équipe de football chez trois bookmakers différents et s'assurer d'un bénéfice certain, d'où la nécessité d'un tel consensus pour les bookmakers.

le bookmaker est le seul à ne pas parier. Il ne veut pas que son profit dépende du résultat du match » (VERGIN et SCRIABIN 1978). Pour ce faire, l'ajustement de la cotation s'avère donc capital pour la rentabilité d'un bookmaker. Lorsque le montant des mises d'un scénario est proportionnellement trop conséquent par rapport aux autres scénarios possibles, le bookmaker va inciter le parieur à miser sur les autres scénarios en augmentant les cotes de ces derniers (et en les rendant donc plus attractives) et en diminuant la cote du scénario pour lequel il y a un nombre important de mises (la rendant moins attractive).

Enfin, la dernière raison impliquant un ajustement de la cotation concerne « les dernières informations en lien avec l'évènement sportif » (TOULEMONDE, 2011). Les bases de données n'étant pas en mesure de pouvoir analyser certains facteurs pouvant influencer l'issue d'une rencontre comme par exemple les blessures, les suspensions, l'état de fraîcheur ou encore la motivation des joueurs, l'intervention humaine se révèle être indispensable. Ainsi le suivi de l'actualité (ou le fameux « news flow » pour les analystes financiers) est pris au sérieux par les bookmakers car toute nouvelle information publique est potentiellement source d'inefficience pour le marché des paris sportifs si elle n'est pas considérée à sa juste valeur.<sup>4</sup>

VERGIN et SCRIABIN (1978) déduisent suite à leur revue « Winning Strategies for Wagering on National Football League Gamesconcluent » que, malgré l'ensemble des moyens qui sont déployés par les bookmakers pour fixer les cotes reflétant les probabilités d'occurrence réelles des différents scénarios d'un évènement, l'unique élément qui est finalement pris en considération par les bookmakers est l'opinion globale des parieurs quant aux probabilités d'occurrence de chacun de ces scénarios. En d'autres termes, « les bookmakers ne proposent pas des cotes reflétant la force réelle des opposants, mais plutôt des cotes reflétant l'opinion des parieurs quant aux forces de ces opposants ».

Puisque la cote est biaisée à la baisse par les bookmakers avec la marge bénéficiaire et que les cotes sont davantage formées par l'agrégation des suppositions des parieurs que par les probabilités statistiques d'occurrence, il apparaît évident que le postulat théorique formulé en (4.3), supposant que les cotes proposées par les bookmakers sont l'inverse des probabilités de réalisation des résultats auxquelles elles sont associées, ne s'applique pas en pratique.<sup>5</sup>

---

<sup>4</sup> Nous étudierons plus en détail ce sujet dans le chapitre 6 de ce mémoire « Influence des « news » sur le marché des paris sportifs »

<sup>5</sup> Nous proposerons, dans le chapitre 6, un modèle ressemblant davantage à la réalité, puisqu'il incorpore le traitement des news dans son calcul de cotation.

## **Les formes alternatives de paris sportifs**

Alors que nous venons d'expliquer en détail le fonctionnement du pari à cote fixe, qui est la forme la plus courante de paris et celle que nous analyserons au cours de ce mémoire, prenons soin de présenter brièvement les deux formes alternatives de paris sportifs. Peu importe la forme, la prise de risque est toujours rémunérée à sa juste valeur. Ce qui change en revanche, d'une forme à l'autre, repose sur la manière de recueillir les enjeux et de les redistribuer.

### Le betting exchange

Le pari à cote fixe est la forme la plus populaire des paris car elle est simple à comprendre. Il suffit de parier sur l'évènement que l'on suppose le plus à même de se produire. Or, très souvent, il est plus facile de prédire ce qui ne va pas se passer. Le betting exchange vient parer à ce problème grâce à sa grande particularité qui propose aux parieurs de pouvoir parier « contre » un évènement. Dans les marchés financiers, on pourrait associer cette possibilité de pari au fait de « shorter » un actif financier. La raison pour laquelle cela est possible est que le betting exchange existe en réalité pour réunir les personnes souhaitant parier sur le scénario d'un évènement sportif avec d'autres personnes qui sont disposés à offrir cette mise. Il n'est dès lors plus question de « bookmaker » puisque les joueurs s'opposent entre eux. Deux options s'offrent à eux ; accepter une mise proposée par un joueur ou fixer une mise en attendant qu'elle soit mise en correspondance avec quelqu'un désirant jouer à cette mise.

Le betting exchange s'apparente ainsi à une véritable « bourse aux paris » et ressemble étrangement aux carnets d'ordres des marchés financiers. De ce fait, on peut dire que le betting exchange est, contrairement au pari à cote fixe (qui, lui, est un marché dirigé par les prix) un marché dirigé par les ordres. Cette forme de pari inclut donc une notion de « trading » qui était encore absente des paris sportifs traditionnels.

### Le pari mutuel

Le pari-mutuel désigne un pari entre joueur où l'ensemble des mises sont « mutualisées » en un pot commun qui est ensuite redistribué aux gagnants en fonction du montant de leurs mises. Alors qu'avec le pari à cote fixe, il est toujours<sup>6</sup> possible de connaître à l'avance le montant de ses gains potentiels, il en est donc tout autre pour le pari-mutuel. En réalité, le

---

<sup>6</sup> A une exception près ; si le parieur décide de miser au « starting price ». Dans ce cas, le parieur accepte de prendre la cote qui sera affichée lorsque l'évènement commence. Cependant, il s'agit d'une possibilité de pari très peu fréquente chez les bookmakers.



montant des gains reçus dépend de trois facteurs distincts : Le montant de la mise du parieur, le montant total des mises et le nombre de personnes ayant placé un pari gagnant. Ainsi, une des spécificités relative au pari mutuel réside sur l'obligation de recevoir un volume important de mises pour offrir un système de cotes intéressant au parieur.

Pour ces différents aspects, cette forme de pari pourrait être comparée au poker, puisqu'il s'agit là aussi de mutualiser des mises et de les redistribuer aux gagnants au prorata de ce qu'ils ont misé, à l'exception près qu'il n'y a, dans une grande majorité des cas, qu'un seul vainqueur au poker.

## **Chapitre 2 - L'efficience informationnelle du marché des paris sportifs**

### **Introduction à l'efficience informationnelle**

Ce deuxième chapitre a pour but de cadrer un peu plus notre sujet de recherche en détaillant de manière plus approfondie l'efficience informationnelle du marché des paris sportifs, tout en présentant la revue littéraire qui en dépend. Précisons d'emblée que, dans ce cas-ci, le mot « information » s'applique principalement aux prix : C'est de la diffusion des cotes dont il s'agit ici. Le terme information aura une toute autre connotation dans le chapitre 6 puisqu'il s'agira à ce moment-là d'informations de type «news».

Le premier à donner une définition précise de l'efficience dans ses travaux est FAMA (1965) qui stipule qu' « un marché est dit efficient si et seulement si l'ensemble des informations disponibles concernant chaque actif coté sur le marché est immédiatement intégré ». JENSEN (1978) viendra apporter, quelques années plus tard, une définition différente mais non-contradictoire à celle de FAMA, en prétendant que « dans un marché efficient, toute prévision dégage un profit nul ». Néanmoins, manquant de précision, la formulation n'a cessé d'être reprise à maintes reprises (notamment par ROLL (1984) qui avançait que « dans un marché efficient, les prix des actifs reflètent toutes les informations pertinentes et fournissent ainsi la meilleure prédiction d'événements futurs, compte tenu des informations actuelles. ») et a débouché en 1999 sur la reformulation suivante qui semble aujourd'hui la plus complète : « Sont réputés efficient, les marchés sur lesquels les prix des actifs cotés intègrent les informations les concernant, de telle manière qu'un investisseur ne puisse, en achetant ou en vendant cet actif, en tirer un profit supérieur aux coûts de transaction engendrés par cette action ». (GILLET, 1999).

Cela signifie qu'en présence d'un marché efficient, même si un investisseur possède une information que les autres agents n'ont pas, il sera incapable d'influencer le marché et de tirer profit à long terme de sa propre stratégie d'investissement. Autrement dit, si le marché est efficient, nul ne peut battre le marché !

Par ailleurs, dans un souci de caractériser au mieux l'efficience relative à chaque marché, FAMA distingue 3 formes d'efficience (faible, semi-forte, forte) qu'il est nécessaire de distinguer ci-dessous pour pouvoir être en mesure de conclure sur l'efficience du marché des paris sportifs de manière plus détaillée.

### **La forme faible de l'efficience**

Un marché sera défini comme étant efficient au sens faible s'il est impossible d'obtenir un profit à long terme en se basant sur les informations passées. Ainsi, si un agent n'est pas en mesure d'exploiter les informations passées d'un actif pour prévoir le prix futur de cet actif, le marché est dit efficient sous sa forme faible.

Dans le cadre des paris sportifs, l'efficience faible désigne dès lors l'impossibilité pour un parieur de se baser sur les cotes et les rendements passés pour prédire les cotes et les rendements futurs. Dans ce contexte, « tester la forme d'efficacité faible revient donc à vérifier s'il est possible de mettre en place des stratégies reposant sur des informations antérieures permettant d'obtenir un rendement positif constant à long terme. Par conséquent, si ces stratégies n'apportent pas un rendement positif significatif et constant, nous pouvons en conclure que le marché des paris sportifs est un marché efficient. » (BARRAUD, 2008)

Nous tenterons dès lors dans la partie analytique de ce mémoire, de voir si les gains relatifs futurs sont totalement, partiellement ou nullement dépendant des gains relatifs passés afin d'être en mesure de conclure sur l'efficience faible de ce marché.

### **La forme semi-forte de l'efficience**

La forme semi-forte de l'efficience d'un marché stipule qu'il est impossible d'obtenir un profit à long terme en se basant sur des informations accessibles au public. Ainsi, on dit d'un marché qu'il est efficient sous sa forme semi-forte si les prix reflètent toute l'information publique disponible. On suppose, par conséquent, un ajustement instantané des prix ; le prix fluctue instantanément à l'annonce d'une information publique.

En finance, ces informations publiques comprennent par exemple les fusions, l'annonce de dividendes, les licenciements, les résultats annuels, les bilans comptables ou encore toutes sortes de variables macro-économiques. Dans le domaine des paris sportifs, par contre, il peut s'agir avant le match, de blessures, de transferts, de suspensions, de changement d'entraîneurs, d'état de forme, de fraîcheur des joueurs ou encore de prévisions météorologiques, et, pendant le match, il peut s'agir de buts marqués, de possession de balle, de cartons jaunes ou rouges, de la qualité du terrain, des conditions météorologiques etc.

Les recherches effectuées dans le cadre de ce mémoire vont donc permettre de savoir si la cote proposée par les bookmakers intègre effectivement ces informations qui sont rendues

publiques. Au-delà de ça, on analysera également la vitesse de l'ajustement puisque, pour que le marché soit efficient au sens semi-fort du terme, il ne doit exister aucun décalage temporel entre l'apparition d'une nouvelle information et le moment où celle-ci est intégrée dans la cotation du bookmaker (ou du moins être si minime, qu'il n'est pas possible pour le parieur d'en tirer profit).

### **La forme forte de l'efficience**

La forme semi-forte de l'efficience prend sens lorsqu'il est impossible d'obtenir un profit à long terme en se basant sur des informations non encore publiées, dites « privées ». Par conséquent, le marché est dit « fortement » efficient si toutes les informations, aussi bien publiques que privées, sont reflétées dans les prix du marché. Si tel est le cas, alors même un investisseur possédant une information que les autres agents ignorent ne seraient pas en mesure de tirer profit de cette information privilégiée. Dans le cadre des paris sportifs, il pourrait par exemple s'agir d'une blessure ou d'une maladie qu'un joueur tente de cacher (et qui n'est donc pas communiquée dans les médias). Dans les cas les plus extrêmes, les informations privées peuvent également comprendre le truchement de matchs ou le dopage de sportifs.

Au vu de la complexité que représente l'étude de cette forme d'efficience, nous n'effectuerons pas d'analyses propres à l'hypothèse d'un marché efficient sous sa forme forte dans le cadre de ce mémoire. En effet, cette forme d'efficience est très difficile à démontrer empiriquement et déboucherait sur un nombre beaucoup trop importants d'hypothèses et/ou de suppositions. De fait, pour prouver l'efficience forte du marché des paris sportifs, il faudrait, dans un premier temps, être en mesure de repérer, après coup, des événements sportifs dans lesquelles certains agents ont détenus des informations privilégiées, et puis, prouver que ces détenteurs d'informations ont effectivement tiré profit de cet avantage pour réaliser un gain, ce qui est difficilement réalisable d'un point de vue pratique.

## **Revue littéraire sur l'efficacité du marché des paris sportifs**

Alors qu'en finance, une revue littéraire évoquant l'efficacité du marché financier et ses trois formes est largement présente et existe depuis de nombreuses années (voir, entre autres BACHELIER (1900), SAMUELSON (1973), MANDELBROT (1971)), les revues scientifiques évoquant l'efficacité du marché des paris sportifs s'avèrent naturellement moins fournies et plus récentes.

La majorité des recherches pour ce marché sont liées à l'analyse de l'efficacité faible et s'avèrent généralement concluantes quant à l'efficacité du marché. Par exemple, POPE et PEEL (1989) ont été de véritables pionniers dans les recherches liées à l'efficacité du marché des paris sportifs en ayant développé un modèle de probabilité linéaire permettant de tester diverses stratégies de paris afin de prouver que le marché est efficace au sens faible du terme. Plus tard, ils ont également réalisé un test sur la forme semi-forte, en se basant sur les prévisions publiées par des spécialistes pour adopter leurs stratégies de paris. Aucune stratégie n'ayant permis d'obtenir des rendements positifs à long termes, ils en ont conclu que le marché était dès lors efficace. WOLFERS et ZITZEWITZ (2004) ont, eux, prouvé que le marché des paris sportifs est capable de regrouper de manière efficace les informations et de fournir des prévisions assez précises des probabilités d'occurrence réelles. CROXSON et READE (2008) arrivent quant à eux à la même conclusion pour le Betting Exchange ; les cotes intègrent les informations pertinentes rapidement et intégralement, ce qui indique un haut niveau d'efficacité.

Néanmoins, certains auteurs n'arrivent pas à cette conclusion. L'inefficacité la plus largement documentée est une anomalie persistante connue sous le nom de « Favourite-longshot bias » (voir entre autres GRIFFITH (1949), QUANDT (1986), GRAY et GRAY (1997), SAUER, 1998, VAUGHAN WILLIAMS (1999, 2005) ou OTTAVIANI et SORENSEN (2008)) que nous avons abordé au cours du chapitre 1.

Notons finalement qu'un nombre important de revues scientifiques se sont intéressées de plus près à la détection de parallèles entre les marchés financiers et le marché des paris sportifs. A titre d'exemple, RUHM (2003) a démontré comment les positions dans certaines options financières pouvaient être assimilées à de simples paris alors que, plus tard, VECER et al. (2006) compareront certains types de paris à des dérivés de crédit.

Nous allons désormais voir si les résultats que l'on obtient sont cohérents avec ceux présents dans la littérature, en menant principalement une analyse sur l'efficacité faible pour le basketball et le baseball, mais en portant également un regard neuf sur l'efficacité semi-forte pour le football, excluant volontairement les tests sur l'efficacité forte qui s'avèrent, comme expliqué précédemment, difficiles à réaliser et à interpréter.

## **Deuxième partie : Tests empiriques et analyses de l'efficience**

### **Chapitre 3 - Méthodologie**

Dans un premier temps, il a fallu identifier des sports dans lesquels le nombre de matchs étaient suffisants que pour pouvoir avoir un échantillon significatif de par sa taille. J'ai donc décidé d'opter pour le baseball et le basketball qui sont tous les deux réputés pour connaître énormément de matchs chaque année.

Le problème rencontré est qu'au plus on augmente la taille d'années étudiées, au plus on a de chance de se voir confronter à des facteurs externes pouvant changer totalement les performances de l'équipe : Nouvel actionnaire/investisseur dans le club, nouveaux joueurs, nouvel entraîneur, nouveaux objectifs, nouvelles mentalités, ...Par contre, si l'on réduit trop le nombre d'années étudiées, les données ne seront pas représentatives car elles résulteront d'une analyse trop ponctuelle de la situation. D'une part, en augmentant le nombre d'années étudiées, je m'exposais donc à avoir un manque de pertinence de par l'influence d'un trop grand nombre de facteurs externes. D'autre part, en diminuant le nombre d'années étudiées, je risquais un manque de cohérence dû à l'étude d'un nombre d'observations trop peu nombreuses.

Pour combler à ce problème, mon parti pris a été de choisir un total de 6 années, ce qui me semble être un compromis intéressant pour avoir des données suffisamment fournies tout en m'assurant qu'elles soient significatives.

Pour chacun des deux sports, 4 équipes de niveau différent ont été sélectionnées. Pour la moitié d'entre elles, j'ai simulé une mise répétitive exclusivement sur la victoire de ces deux équipes et pour l'autre moitié exclusivement sur la défaite.

L'analyse ci-dessous est donc réalisée sur deux sports et huit équipes différentes. Pour quatre d'entre elles (deux dans chaque sport), c'est sur la victoire de tous les matchs de ces équipes que porte l'ensemble des mises alors que pour les quatre autres (deux dans chaque sport), c'est sur la défaite. Cette approche me permettra de voir s'il existe des convergences ou des divergences entre ces deux scénarios et/ou entre ces deux sports. Je commencerai à analyser le basketball, avant de réaliser les mêmes tests pour le baseball. Afin de permettre au lecteur une meilleure compréhension des résultats, je réaliserai, pour la première équipe du premier sport,

une analyse plus approfondie, avant de poursuivre avec une analyse un rien moins détaillée pour les autres équipes puisque le raisonnement est semblable pour l'ensemble des équipes.

### Données recueillies

L'analyse de l'efficience faible a nécessité une recherche préalable de données afin de pouvoir être en mesure de conclure sur la prévisibilité des rendements futurs. Ainsi l'ensemble des données recueillies pour cette analyse ont été récoltées sur le site «Oddsportal»<sup>7</sup>. Les informations suivantes en ont été extraites :

Date : La date de chaque match de l'équipe analysée allant de début 2013 à fin 2018 dans le but de couvrir un total de 6 années.

Match : Le nom des deux équipes qui s'affrontent : l'équipe analysée ainsi que son adversaire du jour. La première équipe mentionnée est l'équipe 1 et évolue à domicile tandis que la seconde équipe est l'équipe 2 et joue à l'extérieur.

Score : Le score final du match opposant l'équipe analysée à son adversaire.

Cote Equipe 1 : La cotation de début de match de l'équipe évoluant à domicile. Cette cotation représente la moyenne des cotations d'une dizaine de bookmakers différents et cette cote est donc d'autant plus significative.

Cote Equipe 2 : La cotation de début de match de l'équipe évoluant à l'extérieur. Cette cotation représente la moyenne des cotations d'une dizaine de bookmakers différents et cette cote est donc d'autant plus significative.

### Données calculées

Sur base de ces données, il m'a été possible de construire et de calculer d'autres variables qui vont s'avérer primordiale pour la réalisation ultérieure de tests empiriques, comme entre autres :

Gain pour 1€ : Montant représentant le bénéfice<sup>8</sup> réalisé sur l'évènement sous l'hypothèse que la mise initiale soit de 1€.

---

<sup>7</sup> <https://www.oddsportal.com/>

<sup>8</sup> Attention que nous parlons bien ici de bénéfice (ou de gain net) et non d'argent reçu. En effet, la mise est déduite de l'argent gagné pour qu'il ne reste au final que le bénéfice engrangé par le pari. Si la mise initiale est de 10€, que le pari est gagnant et que la cotation de l'équipe victorieuse s'élevait à 2.5, le bénéfice ou gain net s'élèvera à 15€ et non 25€.



La formule générale appliquée pour le calcul de ce gain net (G) est, si l'on admet que C(g) représente la cotation de l'équipe gagnante et que M désigne le montant de la mise :

$$\text{En cas de victoire : } G = C(g) - M \quad (3.1)$$

$$\text{En cas de défaite : } G = - M \quad (3.2)$$

Gain relatif : Le gain relatif (GR) représente en pourcentage la proportion du gain net (G) par rapport à la mise de départ (M) et se calcule donc en appliquant la formule suivante :

$$GR = G / M \quad (3.3)$$

Gain cumulé : Le gain cumulé d'un événement donné se définit ici comme étant la somme des gains relatifs de tous les événements passés et présent.

Binaire : Le code binaire correspond au résultat qualitatif de ma mise. Si le pari est gagnant, le gain relatif est positif et le code binaire est égal à 1. Dans le cas où le pari est perdant, le gain relatif est négatif et le code binaire associé à l'évènement est égal à 0.

Nombre de runs : Selon Wald–Wolfowitz, le nombre de runs (R) est calculé en observant l'évolution des codes binaires (0 ou 1 dans ce cas-ci). La formule 3.4 stipule que si le code binaire à l'instant t n'est pas égal au code binaire à l'instant t-1 (on passe donc de 0 à 1 ou de 1 à 0), alors le nombre de runs augmente de 1 unité. En revanche, la formule 3.5 explique, elle, tout le contraire, c'est à dire que si le code binaire à l'instant t est égal au code binaire à l'instant t-1 (on passe de 0 à 0 ou de 1 à 1), alors cela n'augmente pas le nombre de runs de telle sorte que :

$$\text{Si } CB_t \neq CB_{t-1} \text{ alors } R_t = R_{t-1} + 1 \quad (3.4)$$

$$\text{Si } CB_t = CB_{t-1} \text{ alors } R_t = R_{t-1} \quad (3.5)$$

A titre d'exemple, 0 1 0 1 0 représente un total de 5 runs, 0 0 1 1 0 ne compte que 3 runs alors que 0 1 1 0 1 1 0 0 0 1 totalise un nombre de runs égal à 6. Dans son sujet de recherche « *Nonparametric tests for randomness* », WANG (2003), prend lui, par exemple, une autre série binaire (avec des lettres et non des chiffres) en stipulant « qu'une série de données décrivant des hommes et des femmes présentées comme suit M F M F M F M F et M M M M F F F F comprennent respectivement 8 et 2 runs ». Cette partie sera plus largement discutée dans la partie des runs test.

## Bref aperçu et exemples

**Tableau 3.1 :** Aperçu des données recueillies et des données calculées pour les 12 premiers matchs des Houston Astras

N°	Date	Match	Score	Cote Equipe 1	Cote Equipe 2	Gain/Perte pour 1€	Gain relatif	Gain cumulé	Binaire	Nombre de runs
1	samedi 23 février 2013	<a href="#">Philadelphia Phillies - Houston Astros</a>	03:08	<a href="#">1.69</a>	<a href="#">2.19</a>	1,19	119%	119%	1	1
2	lundi 25 février 2013	<a href="#">St.Louis Cardinals - Houston Astros</a>	10:02	<a href="#">1.64</a>	<a href="#">2.28</a>	-1	-100%	19%	0	2
3	mercredi 27 février 2013	<a href="#">Toronto Blue Jays - Houston Astros</a>	01:10	<a href="#">1.61</a>	<a href="#">2.35</a>	1,35	135%	154%	1	3
4	vendredi 1 mars 2013	<a href="#">Houston Astros - St.Louis Cardinals</a>	08:08	<a href="#">2.01</a>	<a href="#">1.80</a>	-1	-100%	54%	0	4
5	samedi 2 mars 2013	<a href="#">Houston Astros - Atlanta Braves</a>	06:05	<a href="#">1.93</a>	<a href="#">1.88</a>	0,93	93%	147%	1	5
6	dimanche 3 mars 2013	<a href="#">Pittsburgh Pirates - Houston Astros</a>	08:06	<a href="#">1.83</a>	<a href="#">1.97</a>	-1	-100%	47%	0	6
7	lundi 4 mars 2013	<a href="#">Detroit Tigers - Houston Astros</a>	08:05	<a href="#">1.62</a>	<a href="#">2.32</a>	-1	-100%	-53%	0	6
8	mardi 5 mars 2013	<a href="#">Washington Nationals - Houston Astros</a>	07:01	<a href="#">1.65</a>	<a href="#">2.26</a>	-1	-100%	-153%	0	6
9	jeudi 7 mars 2013	<a href="#">Houston Astros - Washington Nationals</a>	04:02	<a href="#">2.22</a>	<a href="#">1.68</a>	1,22	122%	-31%	1	7
10	samedi 9 mars 2013	<a href="#">New York Mets - Houston Astros</a>	09:06	<a href="#">1.75</a>	<a href="#">2.09</a>	-1	-100%	-131%	0	8
11	dimanche 10 mars 2013	<a href="#">Houston Astros - Philadelphia Phillies</a>	01:07	<a href="#">2.14</a>	<a href="#">1.72</a>	-1	-100%	-231%	0	8
12	mardi 12 mars 2013	<a href="#">Miami Marlins - Houston Astros</a>	04:09	<a href="#">1.79</a>	<a href="#">2.02</a>	1,02	102%	-129%	1	9

Prenons comme exemple l'ensemble des données pour l'équipe des Houston Astras. Pour cette équipe, j'ai simulé pour tous leurs matchs (de 2013 à 2018) une mise sur la victoire des Houston Astras.

### Exemple 1

Pour le troisième match (le 27 février 2013), on peut voir que les Houston Astras gagnent le match 10 à 1. La cotation de l'équipe gagnante apparait en bleu. Comme j'ai décidé de miser sur la victoire des Houston Astros (voir ci-dessus), le pari est ici gagnant. Si l'on applique la formule (3.1) mentionnée préalablement, mon gain net (G) pour une mise de 1€ est de :

$$G = 2.35\text{€} - 1\text{€}$$

Pour 1€ misé, j'obtiens donc un bénéfice net de 1.35€ et mon gain relatif est donc de

$$\text{Gain relatif} = \text{Gain net} / \text{Mise} = 1.35\text{€} / 1\text{€} = 135\%$$

Mon gain cumulé s'élève lui à 154% que j'obtiens en additionnant les gains relatifs du premier, du deuxième et du troisième événement.

Il s'agit d'un pari gagnant, j'ai un rendement positif, c'est à dire un gain, et mon code binaire est donc de 1.

Mon nombre de runs augmente de 1 car le code binaire de l'évènement précédent diffère du code binaire pour cet évènement ci. On passe ainsi de 2 à 3.

### **Exemple 2**

Pour le septième match des Houston Astras de la saison, on observe une défaite de la part de ces derniers contre les Detroit Diggers pour un résultat final de 8 à 5. De nouveau, on voit apparaître en bleu la cote de l'équipe gagnante. Néanmoins, puisque j'ai parié sur la victoire des Astras, mes pertes s'élèvent à la hauteur de ma mise (voir la formule 3.2). Pour 1€ misé, je perds 1€ et mon gain relatif est de -100% (ou ma perte relative est de 100%). Cette défaite me plonge dans un gain cumulé qui est négatif pour la première fois (-53%). La défaite est par ailleurs synonyme d'un code binaire égal à 0 et puisque le match précédent était également une défaite, le code binaire ne change pas et mon nombre de runs n'augmente pas non plus.

### **Analyse graphique**

Sur base des données présentes ci-dessus, j'ai été en mesure d'effectuer de multiples analyses graphiques me permettant de me prononcer sur la prévisibilité ou non des rendements futurs et donc sur l'efficacité du marché des paris sportifs.

Dans un premier temps, l'évolution graphique des gains relatifs cumulés dans le temps est étudiée. Cette analyse permet d'observer clairement la tendance que prend la courbe des gains relatifs cumulés si l'on mise sur le même évènement (soit la victoire, soit la défaite) d'une même équipe au cours de 6 années.

Deux éléments doivent être considérés pour ce graphique. Le premier est le résultat final à l'échéance. Y a-t-il la présence d'une perte ou d'un gain au bout de ces 6 années ? Une perte prouverait que miser à long terme sur un même évènement pour une même équipe n'est pas profitable pour l'investisseur alors qu'un gain serait le signe qu'il est possible de tirer profit à long terme d'une même stratégie de mise.

La deuxième composante à prendre en considération est la fluctuation de l'évolution graphique. Au plus le graphique démontre une oscillation forte (c'est-à-dire que le gain relatif cumulé ne cesse d'aller à la hausse puis à la baisse), au plus il est possible de supposer que la distribution est aléatoire et qu'il est donc pratiquement impossible de prédire les rendements futurs. Par contre, si nous sommes en présence d'un schéma où on aperçoit des tendances à la

hausse ou à la baisse sur un grand nombre de matchs, et que les fluctuations sont peu élevées, il y a fort à parier que les données observées soient aléatoirement distribuées.

Ensuite, un focus est mis sur l'analyse du gain relatif de chaque match. Ce graphique en nuage de points nous permettra de nous prononcer quant à la présence ou non d'une certaine tendance et apportera une vision plus qualitative de la situation puisque seul le résultat final du pari sera pris en compte (gagnant ou perdant).

Ces deux graphiques, l'un avec une approche quantitative, l'autre avec une optique plutôt qualitative, seront donc très utiles pour avoir une première impression graphique de la prévisibilité des rendements futurs à long terme.

### **Runs Test**

Toujours en partant des données traitées dans mes différents tableaux Excel, il m'a été possible d'effectuer un « *run test* » afin de prouver la dépendance (ou non) entre les rendements passés et les rendements présents et ainsi de conclure quant à l'efficacité faible de cet ensemble de données.

En effet, le run test a pour but de déterminer si une série de données provient d'un processus aléatoire ou non. Nous allons ainsi, grâce à cette procédure statistique, analyser l'occurrence d'événements similaires par rapport à des événements différents.

Pour ce faire, supposons que le run test soit basé sur les hypothèses suivantes :

Hypothèse nulle :  $H_0$  : Les variables sont aléatoirement distribuées (3.6)

Hypothèse alternative :  $H_a$  : Les variables ne sont pas aléatoirement distribuées (3.7)

Dans le cas présent, nous allons procéder en marquant les données par un code binaire car pour appliquer le run test, la variable doit être une variable qualitative.

Le code sera de 1 si la mise débouche sur un gain alors qu'il sera de 0 si la mise mène à une perte (voir la variable « Binaire » des données calculées) Le run test prendra dès lors en compte le signe (rendement négatif ou positif) mais pas la distance (un gain relatif de 500% sera équivalent à un gain relatif de 30%). A ce titre, on peut dire que le run test est complémentaire au test du Khi-Carré qui, lui, prend en compte les distances mais pas les signes, et que l'on aura l'occasion d'étudier plus en profondeur par après.

Concernant le run test, la première étape à effectuer était d'identifier le nombre de paris gagnants et le nombre de paris perdants. Le pari est perdant à partir du moment où le scénario sur lequel la mise a eu lieu n'est pas réalisé. Par conséquent, dans le cas d'un match nul entre deux équipes, le pari sera forcément perdant puisque l'analyse ne comprend que des mises sur la défaite ou la victoire des équipes sélectionnées. J'ai identifié de la sorte :

$$n_1 = \text{Nombre de paris gagnants} \quad (3.8)$$

$$n_2 = \text{Nombre de paris perdants} \quad (3.9)$$

$$\text{avec } n = n_1 + n_2 = \text{Nombre de paris joués au total} \quad (3.10)$$

Ensuite, il m'a fallu calculer le nombre de runs. En appliquant les formules (3.4) et (3.5), il m'a été possible d'identifier le nombre total de runs pour les données observées. Dans le tableau (3.1), la dernière colonne représente le nombre de runs cumulés. Dès lors, en consultant la toute dernière observation (c'est-à-dire le dernier match joué), j'obtiens le nombre de runs que je recherche et que je note :

$$R = \text{Nombre de runs maximum observés pour l'ensemble des données} \quad (3.11)$$

Une fois que l'identification du nombre de runs observés est réalisée, il est nécessaire de calculer le nombre de runs attendus en appliquant la formule suivante :

$$\text{Moyenne : } E(R) = \frac{2n_1n_2}{n} + 1 \quad (3.12)$$

En découle par la suite le calcul de la variance et de son écart-type que j'obtiens de la manière suivante :

$$\text{Variance : } \text{Var} = \frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)} \quad (3.13)$$

$$\text{Ecart type : } \text{StD} = \sqrt{\text{Var}} \quad (3.14)$$

Je décide en outre de fixer mon niveau de signification, c'est-à-dire mon seuil de tolérance à 0.05. Il existe de la sorte un risque de 5% de conclure que mes données ne soient pas aléatoires, alors qu'elles le sont en réalité. J'inscris cela sous la forme suivante :

$$\text{Alpha : } \alpha = 0.05 \quad (3.15)$$

La valeur de mon test statistique sera définie comme étant la différence entre le nombre de runs observés et le nombre de runs attendus ajustée par l'écart type. On aura par conséquent :

$$\text{Test statistique : } Z = \frac{R - E(R)}{StD} \quad (3.16)$$

Enfin, nous constaterons la p-value associée à la valeur Z trouvée préalablement de telle manière que :

$$P\text{-value} = 2P(Z > |z|)^9 \quad (3.17)$$

Une fois la P-value calculée, il sera alors possible de conclure quant à la prévisibilité des rendements futurs. En effet, le run test rejettera l'hypothèse nulle si :

$$P\text{-value} < \alpha$$

C'est-à-dire, en appliquant la formule 3.15, si :

$$P\text{-value} < 0.05 \quad (3.18)$$

### Régression linéaire

La dernière analyse repose sur l'utilisation de régressions linéaires visant à déterminer, une fois de plus, s'il est possible de prévoir les rendements futurs à long terme. La méthode de la régression linéaire va permettre de savoir s'il existe un lien entre les rendements actuels et les rendements passés.

Pour ce faire, nous allons à nouveau étudier nos données de manière quantitative, c'est-à-dire en prenant compte des distances entre les rendements, et de manière qualitative, c'est-à-dire avec pour seul intérêt de savoir s'il s'agissait d'une perte ou d'un gain.

Nous définissons notre première régression linéaire quantitative comme suit :

**Modèle :  $CB_t = \alpha + \beta CB_{t-1} + \epsilon_t$**

**(3.19)**

Où  $CB_t$  = Le rendement à l'instant t sous sa forme binaire (variable à expliquer)

$CB_{t-1}$  = Le rendement à l'instant t-1 sous sa forme binaire (variable explicative)

$\lambda$  = La constante, c'est-à-dire la valeur pour laquelle la droite de régression croise l'axe des Y.

---

<sup>9</sup> Attention nous sommes ici en présence d'un « *two tailed test* » ( test bilatéral ) d'où la présence du multiplicateur 2

$\beta$  = Le coefficient de  $CB_{t-1}$

$\epsilon_t$  = le terme d'erreur (différence entre ce que le modèle prédit et la réalité)

On cherche ici à savoir si  $CB_{t-1}$  a une influence sur  $CB_t$ . Par conséquent, nous allons nous concentrer sur la valeur du coefficient  $\beta$  pour pouvoir conclure de la dépendance ou non entre les deux variables, tout en regardant si les résultats sont significatifs. Le but approuvé de cette régression est de tester la nullité du coefficient  $\beta$ .

Les hypothèses relatives à ce test sont les suivantes :

Hypothèse nulle :  $H_0 : \beta = 0$  : «  $CB_{t-1}$  n'a aucun pouvoir explicatif sur  $CB_t$  »

Hypothèse alternative :  $H_a : \beta \neq 0$  : «  $CB_{t-1}$  a un pouvoir explicatif sur  $CB_t$  »

Ensuite, en ce qui concerne notre deuxième régression linéaire, qui est, elle, plutôt qualitative, nous supposons :

$$\text{Modèle : } GR_t = \lambda + \beta GR_{t-1} + \epsilon_t \quad (3.20)$$

Où  $GR_t$  = Le gain relatif à l'instant  $t$ <sup>10</sup> (variable à expliquer)

$GR_{t-1}$  = Le gain relatif à l'instant  $t-1$  (variable explicative)

$\lambda$  = La constante, c'est-à-dire la valeur pour laquelle la droite de régression croise l'axe des Y.

$\beta$  = Le coefficient de  $GR_{t-1}$

$\epsilon_t$  = Le terme d'erreur, c'est-à-dire la déviation entre ce que le modèle prédit et la réalité.

De même, nous cherchons ici à savoir si  $GR_{t-1}$  a une influence sur  $GR_t$  et c'est donc la variable  $\beta$  qui sera importante à analyser car le but de cette régression est, une nouvelle fois, de tester la nullité de  $\beta$ .

Les hypothèses résultant de ce modèle sont donc :

Hypothèse nulle :  $H_0 : \beta = 0$  : «  $GR_{t-1}$  n'a aucun pouvoir explicatif sur  $GR_t$  »

Hypothèse alternative :  $H_a : \beta \neq 0$  : «  $GR_{t-1}$  a un pouvoir explicatif sur  $GR_t$  »

---

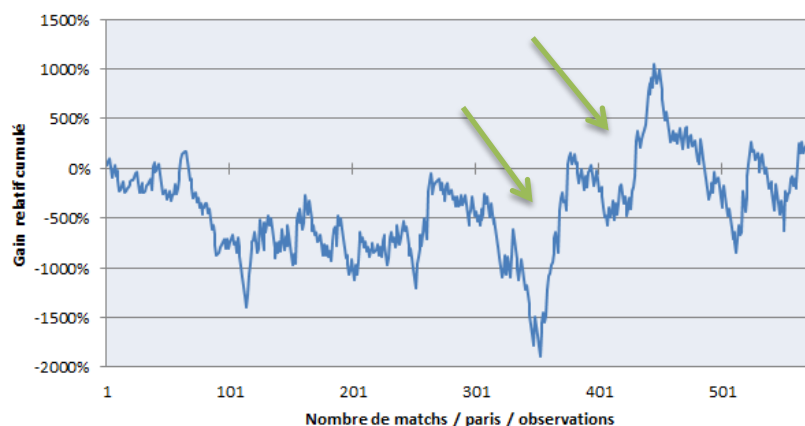
<sup>10</sup> Cette donnée est calculée en utilisant la formule (3.3) et est présente à la huitième colonne du tableau (3.1)

## Chapitre 4 – Test de l’efficience faible pour le basketball

### Analyse du cas n°1 : Mise sur la victoire des Miami Heat

#### Analyse graphique

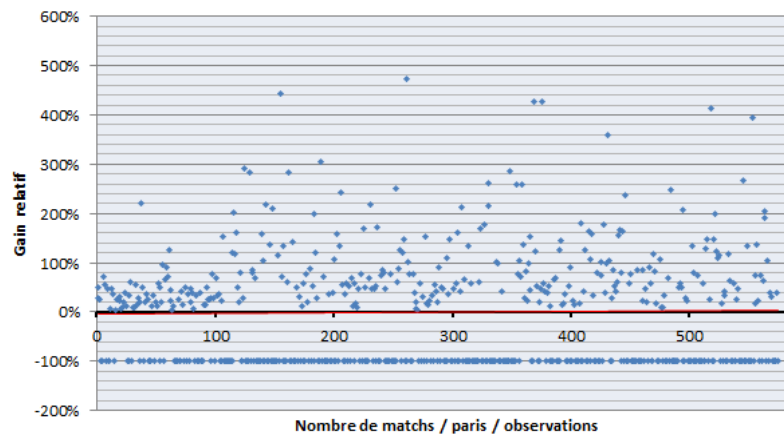
Sur la figure 4.1, nous apercevons globalement une bonne alternance entre gain et perte si ce n’est deux tendances haussières (en vert) qui viennent un peu contraster avec l’oscillation tout à fait normale observée par ailleurs. La perte cumulée après ces 6 années de mises sur la victoire des Miami Heat n’est « que » de (-)200%, ce qui n’est, en réalité pas si mal, même s’il s’agit d’une perte.



*Figure 4.1: Evolution des gains relatifs cumulés entre 2013 et 2018 pour l'équipe des Miami Heat*

La figure 4.2, quant à elle, nous montre que la majorité des gains se trouvent entre 0% et 100%, ce qui est tout à fait normal car l'équipe des Miami Heat est une des plus fortes du championnat américain et les cotes concernant leur victoire est très faible puisque la probabilité de les voir gagner est élevée. L'ensemble des points représentés à -100% comprennent toutes les défaites et les matchs nuls de l'équipe pour lesquels la formule (3.2) est d'application. La ligne rouge désigne la moyenne des gains relatifs au cours des matchs. On voit ici qu'elle part légèrement en dessous de 0% pour finir un rien au-dessus de 0%, signifiant qu'en moyenne, le parieur ne gagne pas d'argent mais n'en perd pas non plus, ce qui est déjà une belle réussite en soi. Ceci va de pair avec l'analyse du premier graphique puisqu'on avait vu que la perte relative n'était que de 200%





*Figure 4.2 : Gain relatif de chaque match entre 2013 et 2018 pour l'équipe des Miami Heat*

### Analyse du Run Test

Le tableau ci-dessous regroupe les résultats obtenus en analysant les données de l'équipe des Miami Heats sur base des formules mentionnées entre les formules (3.8) et (3.17).

*Tableau 4.1 : Réalisation du Run test pour l'équipe des Miami Heat*

n1	n2	n	R	E ( R )	Var	StD	$\alpha$	Z	P-value
307	268	575	298	287,177	142,180	11,924	0,05	0,908	0,364

La première information que révèle le tableau (4.1) est la propension plus importante de victoires (n1) par rapport aux défaites et matchs nuls réunis (n2) pendant les 6 années écoulées. Cela veut donc dire que, même en pariant uniquement sur la victoire d'une équipe qui gagne environ dans 55% des cas comme c'est le cas ici pour les Miami Heat, le rendement final peut-être, comme nous l'avons vu dans l'analyse graphique, négatif !

Le nombre de matchs joués par l'équipe entre 2013 et 2018 s'élève à 575 et forme donc la base de notre analyse. Le nombre de runs observés est quant à lui de 298, ce qui veut dire qu'on passe 298 fois d'un rendement positif (n1) à un rendement négatif (n2) ou inversement au cours de l'ensemble des observations. En appliquant les formules (3.12), (3.13) et (3.14), on obtient respectivement la moyenne, la variance et l'écart type du nombre de runs attendus.

Après avoir calculé la valeur de Z à hauteur de 0.908, il est aisé de trouver une P-value de l'ordre de 0.364. Ainsi la probabilité d'avoir cette série de données si l'hypothèse nulle est vraie est de 36.4%. Par conséquent, comme cette p-value est supérieur au seuil  $\alpha$  de 5%, on ne

peut pas rejeter l'hypothèse nulle<sup>11</sup> de la formule (3.6) et on en conclut donc que les variables sont aléatoirement distribuées au seuil de 5%.

Par ailleurs, la réalisation du run test sur stata vient confirmer l'ensemble des valeurs trouvées via mes formules, comme on le peut le voir sur le tableau ci-dessous

**Tableau 4.2 : Contrôle des résultats via la réalisation du run test sur stata**

```
. runtest Binaire, thresh (0)
N(Binaire <= 0) = 268
N(Binaire > 0) = 307
      obs = 575
      N(runs) = 298
        z = .91
Prob>|z| = .36
```

### Régressions linéaires

Analysons dans un premier temps la régression que l'on avait définie comme étant qualitative et qui vient donc compléter l'analyse du run test puisque dans les deux cas, on ne se soucie pas des « distances » de gain mais juste de savoir s'il y a un rendement positif ou négatif.

**Tableau 4.3 : Régression linéaire de la variable « Binaire » avec sa variable laguée pour l'équipe des Miami Heat**

```
. reg Binaire l1.Binaire
```

Source	SS	df	MS	Number of obs	=	574
Model	.224494235	1	.224494235	F(1, 572)	=	0.90
Residual	142.646586	572	.249382143	Prob > F	=	0.3431
				R-squared	=	0.0016
				Adj R-squared	=	-0.0002
Total	142.87108	573	.249338709	Root MSE	=	.49938

Binaire	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Binaire l1.	-.0396491	.0417892	-0.95	0.343	-.1217281	.0424298
_cons	.5543071	.0305617	18.14	0.000	.4942803	.6143339

La première observation que l'on peut faire dans le tableau 4.3 est que le nombre d'observations a diminué de 1 unité par rapport au tableau 4.1. Ceci s'explique par le fait que, puisque nous utilisons une variable laguée, c'est-à-dire le  $CB_{t-1}$  dans la formule (3.19), il ne peut y avoir de valeur antérieure pour la première donnée.

<sup>11</sup> On ne peut jamais « accepter » une hypothèse nulle.

La constante  $\lambda$  de notre régression linéaire se situe à hauteur de 0.554. Elle est par ailleurs significative puisque sa p-value (0) est inférieure à 0.05. La corrélation entre le code binaire à la période t et le code binaire à l'instant t-1 semble très faible puisqu'elle est de -0.039. Pour s'en convaincre, il est possible d'obtenir le même résultat en demandant directement à stata d'obtenir la corrélation entre les deux variables. On obtient alors :

**Tableau 4.4 : Corrélation entre la variable « binaire » et sa variable laguée pour l'équipe des Miami Heat**

```
. corr Binaire L.Binaire
(obs=574)
```

	Binaire	L. Binaire
Binaire	1.0000	
L1.	-0.0396	1.0000

On remarque de la même occasion, dans le tableau 4.3, que la p-value associée à cette variable indépendante « BinaireL1 » est presque la même que celle obtenue lors de la réalisation du run test ( voir le tableau 4.1) avec une p-value de l'ordre de 0.343. Par ailleurs, la p-value de 0.343 étant supérieur au seuil  $\alpha$  de 0.05%, on en déduira qu'on ne peut pas rejeter l'hypothèse nulle du modèle 3.19 et que par conséquent, les variables sont bien aléatoirement distribuées au seuil de 5%. En analysant la t-stat (qui est comprise entre -1.96 et 1.96) ou l'intervalle de confiance (qui comprend la valeur 0), il nous est possible de tirer une conclusion semblable à celle-ci.

Par ailleurs, la somme expliquée des carrés (0.224) est infime par rapport à la somme résiduelle des carrés (142.646). Ceci explique donc que le  $R^2$  soit très faible puisqu'il est défini comme étant le rapport entre la première et la seconde. Il en résulte que notre modèle n'explique pratiquement rien et qu'il soit ainsi qualifié de « nul ». En effet, dans ce cas-ci, seulement 0.16% des changements de la variable binaire CB sont expliqués par sa variable binaire laguée  $CB_{t-1}$ . En outre, la F-Stat, qui est égale au Khi-carré de Pearson comme nous pouvons le voir dans le tableau ci-dessous, très basse (0.90) vient confirmer cette impression.

**Tableau 4.5 : Test du Khi-carré entre la variable « binaire » et sa variable laguée pour l'équipe des Miami Heat**

```
. tabulate Binaire lBinaire, chi2
```

Binaire	lBinaire		Total
	0	1	
0	119	149	268
1	148	158	306
Total	267	307	574

Pearson chi2(1) = 0.9019 Pr = 0.342

Le  $R^2$  corrigé est logiquement encore plus proche de 0 puisqu'il prend en compte la perte de degré de liberté lié à l'unique variable explicative incluse dans le modèle.

Ainsi, après avoir analysé la régression qualitative, considérons de plus près la régression qui elle, prend en compte l'ampleur des gains ou des pertes dans son analyse.

**Tableau 4.6 : Régression linéaire de la variable « Gainrelatif » avec sa variable laguée pour l'équipe des Miami Heat**

```
. reg Gainrelatif l.Gainrelatif
```

Source	SS	df	MS	Number of obs	=	574
Model	1.86195624	1	1.86195624	F(1, 572)	=	1.51
Residual	703.184047	572	1.22934274	Prob > F	=	0.2189
				R-squared	=	0.0026
				Adj R-squared	=	0.0009
Total	705.046003	573	1.23044678	Root MSE	=	1.1088

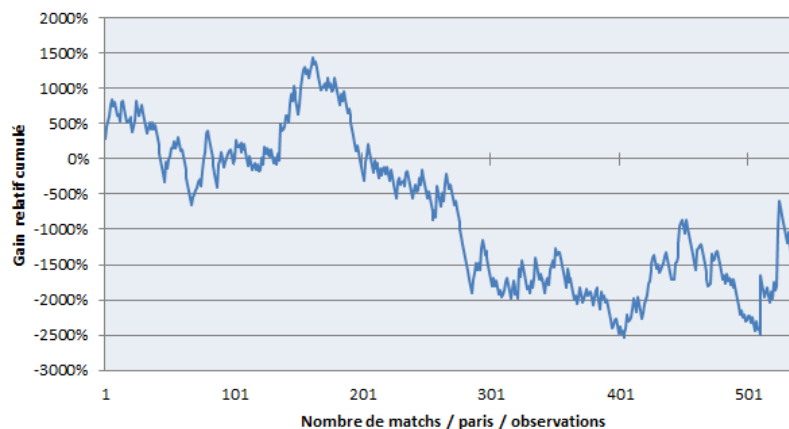
Gainrelatif	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Gainrelatif L1.	-.0514232	.0417841	-1.23	0.219	-.1334921	.0306457
_cons	-.0046121	.0462787	-0.10	0.921	-.0955091	.0862849

Le nombre d'observations reste bien évidemment inchangé (574), la constante est presque nulle (-0.004) alors que le coefficient de la variable  $GR_{t-1}$  n'est que de -0.051. Puisque la p-value qui lui est associée est supérieure à la valeur critique  $\alpha$ , on ne peut pas rejeter l'hypothèse nulle, on conclut que la différence n'est pas significative au seuil de 5% et ainsi que la variable  $GR_{t-1}$  n'a aucun pouvoir explicatif sur la variable  $GR_t$ .

## Analyse du cas n°2 : Mise sur la victoire des New Orleans Pelicans

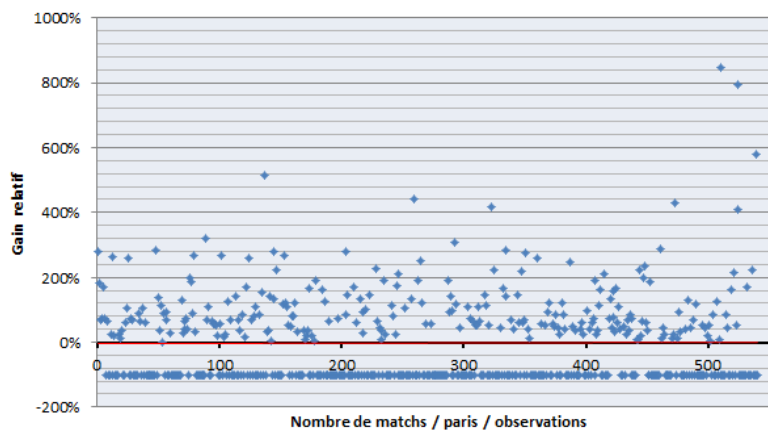
### Analyse graphique

La figure 4.3 met en évidence une évolution que l'on peut qualifier de classique puisque l'on peut clairement observer une succession de gains et de pertes sans observer de tendance à proprement parler. Par ailleurs, le gain relatif cumulé affiche un total de près de -1000% à la fin de l'année 2018.



*Figure 4.3 : Evolution des gains relatifs cumulés entre 2013 et 2018 pour l'équipe des New Orleans Pelicans*

Sur la figure 4.4, on peut constater une fois de plus que les gains relatifs positifs tournent principalement entre 0% et 200%, mais contrairement à l'équipe des Miami Heat, les New Orleans Pelicans ont eu à deux reprises un gain relatif de près de 800%, ce qui peut s'expliquer par le fait que cette équipe soit perçue comme étant moins bonne que les Miami Heats et que la cote associée à la victoire des New Orleans Pelicans soit par conséquent quelques fois très élevée.



*Figure 4.4 : Gain relatif de chaque match entre 2013 et 2018 pour l'équipe des New Orleans Pelicans*

## Analyse du Run test

**Tableau 4.7 : Réalisation du Run test pour l'équipe des New Orleans Pelicans**

n1	n2	n	R	E ( R )	Var	StD	$\alpha$	Z	P-value
250	291	541	264	269,946	133,450	11,552	0,05	0,515	0,607

Le run-test effectué pour les New Orleans Pelicans prend en compte les 541 matches joués par l'équipe lors des 6 dernières années. On constate que la p-value (0.607) est supérieure au seuil  $\alpha$  de 0.05, ce qui signifie qu'il n'est pas possible de rejeter l'hypothèse nulle stipulant que les variables sont aléatoirement distribuées. On conclut ainsi que cette série de données est distribuée de manière aléatoire.

## Régression linéaire

**Tableau 4.8 : Régression linéaire de la variable « binaire » avec sa variable laguée pour l'équipe des New Orleans Pelicans**

**reg Binaire l.Binaire**

Source	SS	df	MS	Number of obs	=	540
Model	.055195402	1	.055195402	F(1, 538)	=	0.22
Residual	134.128138	538	.249308807	Prob > F	=	0.6382
				R-squared	=	0.0004
				Adj R-squared	=	-0.0014
Total	134.183333	539	.24894867	Root MSE	=	.49931

Binaire	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Binaire L1.	.0202759	.043092	0.47	0.638	-.0643733 .1049251
_cons	.4517241	.0293204	15.41	0.000	.3941276 .5093206

Le coefficient de  $CB_{t-1}$  est proche de 0 puisqu'il ne s'élève qu'à hauteur de 0.02, ce qui veut déjà dire que l'influence de  $CB_{t-1}$  sur  $CB_t$  est très faible voir insignifiante. La p-value qui lui est associée (0.638) est de nouveau fort proche de la p-value obtenue lors du run-test, ce qui démontre une certaine cohérence entre le run test et la régression linéaire qualitative. Ainsi, puisque la p-value est supérieure au seuil alpha, il n'est pas possible de rejeter l'hypothèse nulle et on décrète ainsi que  $CB_{t-1}$  n'a aucun pouvoir explicatif sur  $CB_t$ . En outre, il est possible de constater que le coefficient de détermination  $R^2$  est nul, ce qui signifie que la qualité de notre prédiction est mauvaise. Si l'on décide maintenant de régresser les données en prenant en compte l'ampleur des gains et des pertes, on obtient le tableau suivant :

**Tableau 4.9 : Régression linéaire de la variable « Gainrelatif » avec sa variable laguée pour l'équipe des New Orleans Pelicans**

`. reg Gainrelatif L1.Gainrelatif`

Source	SS	df	MS	Number of obs	=	540
Model	.924110964	1	.924110964	F(1, 538)	=	0.54
Residual	914.83722	538	1.70044093	Prob > F	=	0.4613
				R-squared	=	0.0010
				Adj R-squared	=	-0.0008
Total	915.761331	539	1.69900061	Root MSE	=	1.304

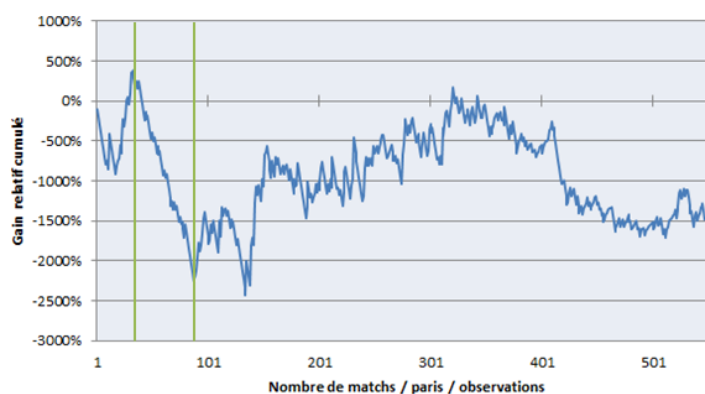
Gainrelatif	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Gainrelatif L1.	-.0316481	.0429306	-0.74	0.461	-.1159803	.052684
_cons	-.0225091	.0561193	-0.40	0.689	-.132749	.0877308

Celui-ci conduit aux mêmes conclusions que pour la régression avec la variable « binaire » ; un modèle qui n'explique pas grand-chose avec des résidus très conséquents, un coefficient  $\beta$  proche de 0 qui n'est, en outre, pas significatif puisque la p-value est supérieur à 0.05 et donc  $R_{t-1}$  n'a aucun pouvoir explicatif sur  $R_t$

### Analyse du cas n°3: Mise sur la défaite des Chicago Bulls

#### Analyse graphique

La figure ci-dessous, laisse supposer une alternance normale dans l'évolution du gain relatif cumulé, si ce n'est une année 2013 qui sort un peu du lot. En effet, les Chicago Bulls développent à la fin de cette année là un jeu d'une grande qualité leur permettant de gagner 35 matchs sur 50 (entre les barres rouges), ce qui s'avère être une catastrophe pour le parieur ayant misé sur la défaite des Chicago Bulls, voyant son gain relatif cumulé passer de +381% à -2243% en l'espace de peu de temps. Au final, la perte de ce parieur se chiffrera à 1287%.

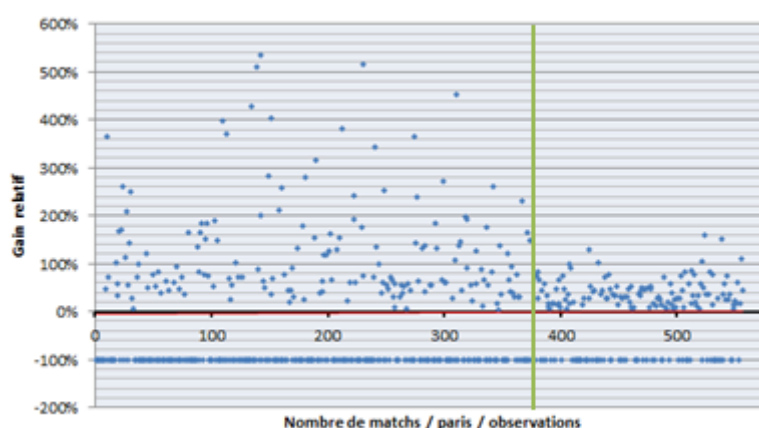


**Figure 4.5 : Evolution des gains relatifs cumulés entre 2013 et 2018 pour l'équipe des Chicago Bulls**



La figure 4.6 est quant à elle, particulière. En effet, on peut apercevoir que pour les 370 premiers matchs (à gauche de la ligne verte), c'est-à-dire de 2013 à 2016, les gains relatifs sont assez dispersés et plutôt peu nombreux. Ceci s'explique par le fait que puisque nous sommes ici dans une situation où l'ensemble des mises se fait sur la défaite de l'équipe, si l'équipe est compétitive, les cotes associées à la défaite de l'équipe sont élevées. Comme les Chicago Bulls réalisent de 2013 à 2016 de très bons matchs et sont donc compétitifs, parier sur la défaite de cette équipe est peu fréquent mais très rentable (parfois jusqu'à 600% de gain relatif sur une mise).

Par contre le contraste avec les deux dernières années (après la ligne verte) est saisissant. En effet, on voit que les gains relatifs se multiplient à partir de 2017 et jusque 2018 mais surtout que les gains sont moins importants. On se trouve donc dans une situation totalement opposée par rapport aux 4 premières années où les Chicago Bulls ont plus tendance à perdre, ce qui veut dire que les gains sont plus nombreux mais moins profitables puisque les cotes associées à leur défaite diminuent fortement.



*Figure 4.6 : Gain relatif de chaque match entre 2013 et 2018 pour l'équipe des Chicago Bulls*

### Analyse du Run test

*Tableau 4.10 : Réalisation du Run test pour l'équipe des Chicago Bulls*

n1	n2	n	R	E ( R )	Var	StD	$\alpha$	Z	P-value
289	267	556	270	278,565	138,315	11,761	0,05	0,728	0,467

A nouveau, comme c'était le cas pour les Miami Heat, on observe que le nombre de paris gagnants est supérieure au nombre de pari perdants. Par contre, dans le cadre d'une analyse où l'on simule l'ensemble des mises sur la défaite de l'équipe comme c'est le cas ici, cela signifie que l'équipe a plus perdu qu'elle n'a gagné ou fait match nul. Pour le parieur,



c'est évidemment une bonne chose. Néanmoins, on peut constater une fois de plus que nonobstant le nombre plus importants de paris gagnants que perdants, les rendements cumulés sur les 6 années sont eux négatifs (voir figure 4.5).

La p-value résultant du run test s'élève à 0.467. De ce fait, il y a 36.4% de chance d'obtenir cette séquence de données si l'hypothèse nulle est vraie. Puisque ce chiffre est supérieur au à 5%, qui est le seuil que l'on a préalablement défini, on décide de ne pas rejeter l'hypothèse nulle et ainsi de conclure que ces données sont aléatoirement distribuées.

### Régression linéaire

Comme nous pouvons l'observer, la proximité entre la p-valeur du run test et la p-valeur de la régression avec les variables qualitatives est encore une fois confirmée. Le tableau 4.10 met en évidence un modèle qui n'explique que très peu de choses, et ceci est dû à un R-squared qui s'avère être très faible (0.0008). Le coefficient de notre régression est de 0.029, ce qui annonce une relation presque nulle entre la variable dépendante et indépendante, mais il est de toute façon non-significatif puisque la p-value (0.494) est supérieure à  $\alpha$ . On peut donc conclure que la variable Binaire au temps t-1 n'a aucun pouvoir explicatif sur la variable Binaire au temps t.

**Tableau 4.11 : Régression linéaire de la variable « binaire » avec sa variable laguée pour l'équipe des Chicago Bulls**

**. reg Binaire l1.Binaire**

Source	SS	df	MS	Number of obs	=	555
Model	.1173609	1	.1173609	F(1, 553)	=	0.47
Residual	138.394351	553	.250261032	Prob > F	=	0.4938
				R-squared	=	0.0008
				Adj R-squared	=	-0.0010
Total	138.511712	554	.25002114	Root MSE	=	.50026

Binaire	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Binaire l1.	.0291042	.0425002	0.68	0.494	-.0543774 .1125859
_cons	.505618	.0306155	16.52	0.000	.4454811 .5657548

Le tableau 4.12, illustrant la régression entre le gain relatif actuel et son gain relatif antérieur, démontre un coefficient de -0.033, ce qui est très bas. Puisque, entre autre, l'intervalle de confiance comprend la valeur 0, on va décider de ne pas rejeter l'hypothèse nulle et d'admettre que le gain relatif antérieur n'a pas de pouvoir explicatif sur le gain relatif actuel. Il est également important de noter que la variable  $R^2$  est égale à 0.0008, ce qui signifie que la

variable explicative n'est responsable qu'à hauteur de 0.08% des changements de la variable dépendante.

**Tableau 4.12 : Régression linéaire de la variable « Gainrelatif » avec sa variable laguée pour l'équipe des Chicago Bulls**

```
. reg Gainrelatif l1.Gainrelatif
```

Source	SS	df	MS	Number of obs	=	555
Model	.828752039	1	.828752039	F(1, 553)	=	0.62
Residual	741.63568	553	1.34111334	Prob > F	=	0.4321
				R-squared	=	0.0011
				Adj R-squared	=	-0.0007
Total	742.464432	554	1.3401885	Root MSE	=	1.1581

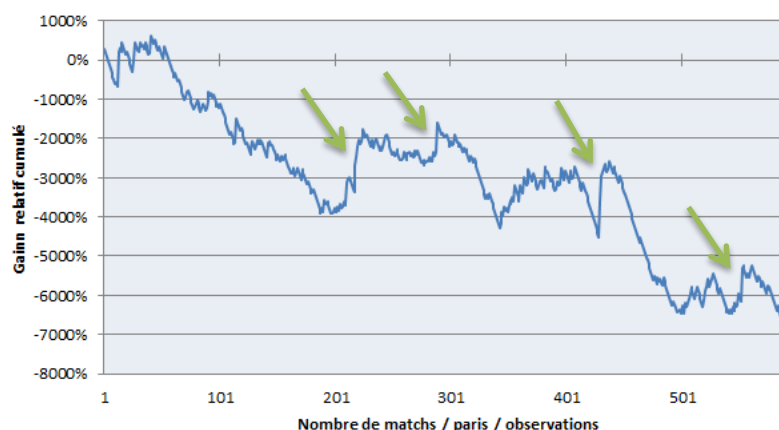
  

Gainrelatif	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Gainrelatif L1.	-.033393	.0424791	-0.79	0.432	-.1168332	.0500472
_cons	-.0221876	.0491676	-0.45	0.652	-.1187658	.0743906

#### Analyse du cas n°4: Mise sur la défaite des Houston Rockets

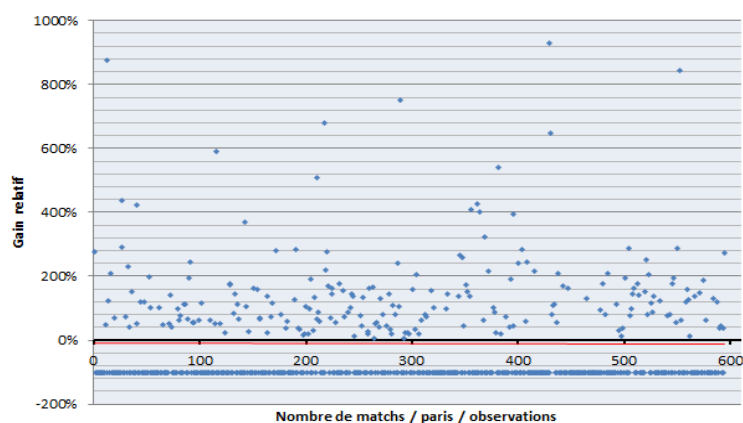
##### Analyse graphique

Le graphique ci-dessous met en exergue une particularité qu'il convient d'épingler ici. En vert, sont représentés des tendances haussières qui, au vue de l'échelle qui est reprise pour ce graphique, sont d'une ampleur importante. Celles-ci résultent, comme nous allons le voir, d'un seul et unique match, c'est-à-dire d'une seule et unique mise. Le rendement relatif cumulé affiche quant à lui un total à l'échéance de -6348%.



**Figure 4.7: Evolution des gains relatifs cumulés entre 2013 et 2018 (Houston Rockets)**

Avec un des meilleurs joueurs au monde des dans leurs rangs, en la personne de James Harden, l'équipe des Houston Rockets fait chaque année office de favori ou en tout cas d'outsider pour le titre. La figure 4.8 nous montre, non seulement que les gains relatifs positifs sont peu nombreux mais qu'ils sont d'une ampleur considérable. En effet, puisque l'équipe est très forte, ils perdent peu de matchs mais les cotes associées à la défaite de leur équipe sont très élevées. Tellement élevées que le gain relatif est parfois de près de 1000%, soit une cote de 11 ! Ainsi, la figure 4.7 illustre à merveille que le gain relatif cumulé peut remonter de manière spectaculaire, et ce grâce à une seule mise, alors qu'il ne peut au maximum descendre que de 100% par match. Nous sommes donc dans le cas précis où la mise est très risquée car sa probabilité d'occurrence est faible, mais où le rendement peut être très élevé.



*Figure 4.8 : Gain relatif de chaque match entre 2013 et 2018 (Houston Rockets)*

### Analyse du Run test

*Tableau 4.13 : Réalisation du Run test pour l'équipe des Houston Rockets*

n1	n2	n	R	E ( R )	Var	StD	$\alpha$	Z	P-value
214	381	595	275	275,064	125,988	11,224	0,05	0,006	0,995

Comme l'on pouvait s'y attendre au vue de ce qui a été dit plus haut, le nombre de paris gagnants (214) est très faible par rapport aux paris perdants (381). Seuls 36% des matchs ont été perdus par les Houston Rockets et ont donc rapporté de l'argent au parieur ayant misé exclusivement sur leur défaite.

La singularité de cette série de données continue puisque l'on peut voir que le nombre de runs observés (275) est pratiquement égal au nombre de runs attendus (275.064). Cela a pour conséquence que la valeur de Z soit très faible (voir formule 3.16) et que la p-value soit quant à elle très grande. Avec une p-value de 0.995, il y a en effet 99.5% de chance d'obtenir cette

séquence de données si l'hypothèse nulle est vraie et on va donc naturellement dire que ces données sont aléatoirement distribuées au seuil de 5%.

### Régression linéaire

Concernant la première régression linéaire, ce que l'on a cessé de constater depuis lors ne changera pas. Le coefficient  $\beta$  reste très proche de 0, la p-value, presque semblable à celle du run test, atteste de la non-significativité des données et le modèle n'explique rien avec un R-squared qui demeure nul.

**Tableau 4.14: Régression linéaire de la variable « binaire » avec sa variable laguée pour l'équipe des Houston Rockets**

. reg Binaire l.Binaire

Source	SS	df	MS	Number of obs	=	594
Model	.001050205	1	.001050205	F(1, 592)	=	0.00
Residual	136.620162	592	.230777301	Prob > F	=	0.9462
				R-squared	=	0.0000
				Adj R-squared	=	-0.0017
Total	136.621212	593	.230389902	Root MSE	=	.48039

Binaire	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Binaire l1.	-.0027725	.0410996	-0.07	0.946	-.0834913 .0779462
_cons	.3595801	.0246113	14.61	0.000	.311244 .4079161

La deuxième régression, elle, met en évidence un coefficient  $\beta$  légèrement plus élevé mais toujours peu représentatif avec seulement 0.036. L'intervalle de confiance comprenant la valeur 0, la p-value étant supérieur à  $\alpha$  et la t-stat comprise entre -1.96 et 1.96 sont autant d'éléments qui nous permettent, ici aussi, de ne pas rejeter l'hypothèse nulle et ainsi d'en déduire que la variable explicative n'a aucun pouvoir explicatif sur la variable à expliquer.

**Tableau 4.15 : Régression linéaire de la variable « Gainrelatif » avec sa variable laguée pour l'équipe des Houston Rockets**

. reg Gainrelatif l.Gainrelatif

Source	SS	df	MS	Number of obs	=	594
Model	1.74695793	1	1.74695793	F(1, 592)	=	0.79
Residual	1311.22862	592	2.2149132	Prob > F	=	0.3748
				R-squared	=	0.0013
				Adj R-squared	=	-0.0004
Total	1312.97558	593	2.21412407	Root MSE	=	1.4883

Gainrelatif	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Gainrelatif l1.	.0364725	.0410679	0.89	0.375	-.044184 .117129
_cons	-.1074834	.0612353	-1.76	0.080	-.2277483 .0127815

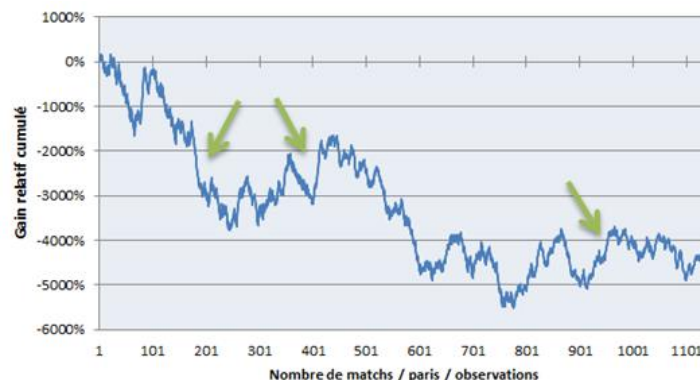
## Chapitre 5 – Test de l’efficience faible pour le baseball

Passons désormais à une analyse semblable mais pour un autre sport : le baseball. Le nombre de données analysées, c’est-à-dire le nombre de mises, va presque doubler ici puisqu’au baseball, il y a presque 1 match tous les deux jours ce qui représente plus de 1000 matchs en 6 ans.

### Analyse du cas n°1 : Mise sur la victoire des Houston Astras

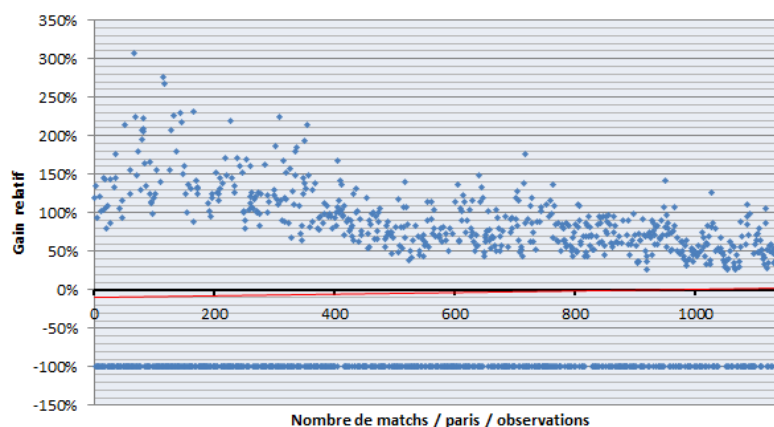
#### Analyse graphique

Sur la figure 5.1, il est possible de constater que quelques tendances durent dans le temps. Par exemple, la première flèche verte présente sur le graphique représente 15 défaites consécutives pour l’équipe des Houston Astras. Au-delà d’être très néfaste pour le parieur ayant misé sur la victoire répétitive de cette équipe (puisque c’est une chute de 1500% du gain relatif cumulé en 15 matchs, soit la perte maximum envisageable), c’est la preuve que les Houston Astras n’ont cessé d’enchaîner le très mauvais, mais aussi le très bon. Car le cas contraire existe aussi, avec la seconde flèche verte, qui elle, illustre le fait que cette même équipe ait enchaîné 10 victoires consécutives en 2015. De nombreux autres cas, plus discrets, existent également. A titre d’exemple, la troisième flèche montre une tendance haussière qui fait passer le gain relatif cumulé de -4725% à -4035%. On pourrait se dire que cela ne représente qu’une demi-douzaine de victoires d’affilés, or, les cotes étaient tellement basses, que cela représente en fait un total de 11 victoires d’affilés, c’est-à-dire encore plus que pour la deuxième flèche ! Ainsi, il est important de ne pas confondre l’ampleur des courbes avec l’ampleur des séries d’évènements consécutifs. Un regain de 1000% concernant le gain relatif cumulé peut aussi bien représenter 1 match (avec une très grosse cote) qu’une vingtaine de matchs consécutifs (avec de faibles cotes).



*Figure 5.1: Evolution des gains relatifs cumulés entre 2013 et 2018 (Houston Astras)*

La figure suivante vient corroborer ces propos. On aperçoit que les gains relatifs positifs étaient plus élevés mais moins fréquents dans la première partie des observations alors qu'ils deviennent plus nombreux mais également plus bas lors de la seconde partie. Cela s'explique par le fait que l'équipe a changé de statut au cours de ces 6 années, passant d'outsider à véritable favori.



*Figure 5.2 : Gain relatif de chaque match entre 2013 et 2018 (Houston Astros)*

### Run test

Les résultats singuliers obtenus lors de la réalisation du run test pour cette équipe n'ont rien d'étonnants. Comme on l'a vu, un nombre important de victoires ou de défaites consécutives ont été repéré graphiquement. Le nombre de runs observés (542) est ainsi largement en dessous du nombre de runs attendus (573.374). Ceci a pour conséquence une p-value qui s'écroule totalement par rapport aux autres cas, pour arriver à 0.063. Néanmoins, la conclusion demeure la même puisque cette valeur est toujours supérieur à notre seuil alpha de 5% ; on ne rejette pas  $H_0$  et on déclare que les variables sont distribuées de manière aléatoire.

A un seuil de 10% par contre, on aurait dû en déduire tout autre chose. En effet, s'il on avait défini le seuil de signification à 10% au lieu de 5%, on aurait dû, dans ce cas-ci, rejeter l'hypothèse nulle et ainsi affirmer que les variables ne sont pas aléatoirement distribuées.

*Tableau 5.1 : Réalisation du Run test pour l'équipe des Houston Astros*

n1	n2	n	R	E ( R )	Var	StD	$\alpha$	Z	P-value
581	564	1145	542	573,374	285,874	16,908	0,05	1,856	0,063

## Régression linéaire

Lorsque l'on regarde maintenant le tableau 5.2, on se rend compte qu'il va, une fois de plus, de pair avec le run test. En effet, la p-value associée à la variable « BinaireL1 » est aussi basse que pour le run test (0.068). Par ailleurs, on constate également que la t-stat est presque supérieure à 1.96 où que l'intervalle de confiance comprend tout juste la valeur 0, ce qui vient confirmer le fait que nous soyons ici dans le cas où le seuil est presque franchi. Néanmoins, puisque 0.068 est supérieur au seuil de 0.05, on décidera de ne pas rejeter l'hypothèse nulle et on soulignera de la sorte que CBt-1 n'a aucun pouvoir explicatif sur CBt, d'autant plus que le coefficient est de toute façon insignifiant (0.053) et que la variable explicative  $R^2$  est presque nulle.

**Tableau 5.2 : Régression linéaire de la variable « binaire » avec sa variable laguée pour l'équipe des Houston Astros**

. reg Binaire 1.Binaire

Source	SS	df	MS	Number of obs	=	1,144
Model	.833433296	1	.833433296	F(1, 1142)	=	3.34
Residual	285.110623	1,142	.249659039	Prob > F	=	0.0679
				R-squared	=	0.0029
				Adj R-squared	=	0.0020
Total	285.944056	1,143	.250169778	Root MSE	=	.49966

Binaire	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Binaire L1.	.0539891	.0295491	1.83	0.068	-.0039875	.1119657
_cons	.4795737	.0210581	22.77	0.000	.4382568	.5208906

Si l'on prend en compte l'ampleur des gains et des pertes, on arrive à une régression relativement différente. De fait, le tableau 5.3 dévoile un coefficient plus faible avec une p-value plus élevée et un R-squared encore moins important. On retrouve donc des données un peu plus standards par rapport à ce que l'on a pu voir auparavant. Dans le cadre de cette régression, on attestera clairement que le rendement à l'instant t-1 n'a aucune influence sur le rendement à l'instant t puisqu'une fois de plus, on ne peut pas rejeter l'hypothèse nulle.



**Tableau 5.3 : Régression linéaire de la variable « Gainrelatif » avec sa variable laguée pour l'équipe des Houston Astros**

**. reg Gainrelatif l1.Gainrelatif**

Source	SS	df	MS	Number of obs	=	1,144
Model	.117110753	1	.117110753	F(1, 1142)	=	0.12
Residual	1128.36183	1,142	.988057641	Prob > F	=	0.7307
				R-squared	=	0.0001
				Adj R-squared	=	-0.0008
Total	1128.47894	1,143	.987295657	Root MSE	=	.99401

Gainrelatif	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Gainrelatif l1.	.0101845	.0295823	0.34	0.731	-.0478573 .0682263
_cons	-.0387036	.0294091	-1.32	0.188	-.0964056 .0189983

## Analyse du cas n°2 : Mise sur la victoire des Philadelphia Philies

### Analyse graphique

Le graphique ci-dessous a la spécificité de ne jamais présenter un gain relatif cumulé positif. On le voit, la courbe ne cesse de chuter au fur et à mesure des années pour arriver -9511%. Les gains relatifs cumulés semblent décroître de manière constante, ce qui démontre que la part des défaites par rapport aux victoires a été conséquente mais aussi constante durant les 6 années analysées



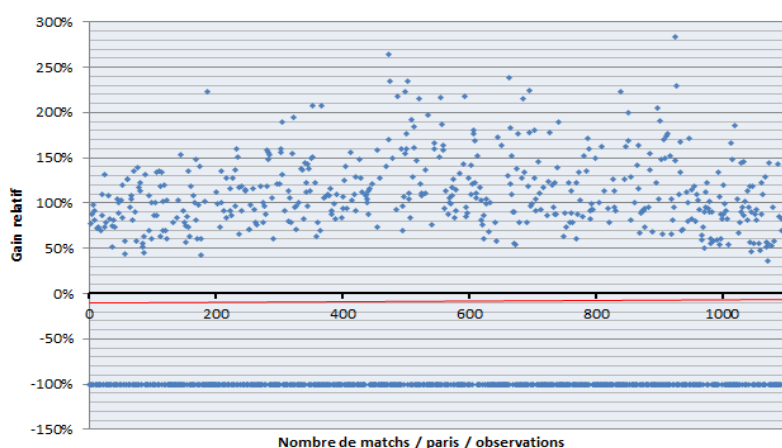
**Figure 5.3 : Evolution des gains relatifs cumulés entre 2013 et 2018 (Philadelphia Philies)**

Le constat effectué sur la figure 5.3 est encore plus visible sur la figure 5.4. La proportion de matchs ayant amené un gain relatif positif est dérisoire par rapport au nombre de matchs ayant occasionné une perte. En outre, les cotes proposées par les bookmakers pour la victoire des



Philadelphia Philies ne permettent pas de compenser le nombre important de matchs nuls ou de matchs perdus puisque le gain relatif positif maximum n'atteint même pas 300%.

On est dans une situation où l'équipe est perçue comme étant relativement bonne, mais qui perd une grosse partie de ses matchs. Pour un parieur décidant de ne miser que les victoires d'une même équipe, il s'agit d'une des pires situations.



*Figure 5.4 : Gain relatif de chaque match entre 2013 et 2018 pour l'équipe des Philadelphia Philies*

### Run test

On l'a vu, le gain relatif cumulé est fortement négatif et ceci pourrait être dû à la propension de défaites ou de matchs nuls (636) beaucoup plus importants par rapport aux victoires (486).

Ce qui est le plus marquant, ceci dit, c'est la grande proximité entre le nombre de runs attendus (551.973) et le nombre de runs observés (552). Puisque l'estimation est presque identique à ce qu'on observe dans les faits, la valeur de la Z-stat est infime, ce qui engendre une p-value proche de 1. Ainsi, on peut sans hésiter décider de ne pas rejeter l'hypothèse nulle puisque 0.998 est nettement supérieur à 0.05. On en déduit par conséquent que les variables sont aléatoirement distribuées.

*Tableau 5.4 : Réalisation du Run test pour l'équipe des Philadelphia Philies*

n1	n2	n	R	E ( R )	Var	StD	$\alpha$	Z	P-value
486	636	1122	552	551,973	270,313	16,441	0,05	0,002	0,998

## Régression linéaire

Le coefficient de la variable indépendante « BinaireL1 » est nul et n'est pas significatif au seuil de 5%. On ne peut dès lors pas rejeter l'hypothèse nulle et on décrète par conséquent que la variable explicative  $CB_{t-1}$  n'a aucune influence sur la variable à expliquer  $CB_t$ .

En outre, le R-squared est également égal à 0 ce qui désigne qu'on a un modèle qui n'explique rien puisque les résidus représentent presque 100% de celui-ci.

**Tableau 5.5 : Régression linéaire de la variable « binaire » avec sa variable laguée pour l'équipe des Philadelphia Philies**

`. reg Binaire L1.Binaire`

Source	SS	df	MS	Number of obs	=	1,121
Model	.000260278	1	.000260278	F(1, 1119)	=	0.00
Residual	275.29858	1,119	.246021966	Prob > F	=	0.9741
				R-squared	=	0.0000
				Adj R-squared	=	-0.0009
Total	275.29884	1,120	.245802536	Root MSE	=	.49601

Binaire	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Binaire L1.	-.0009726	.0299013	-0.03	0.974	-.0596415 .0576964
_cons	.4339623	.0196679	22.06	0.000	.3953721 .4725524

C'est la première fois que la seconde régression semble coïncider avec la première et ce, en tous points ! En effet, autant le R-squared, que le coefficient  $\beta$  en passant par sa p-value ressemblent presque au centième près à la précédente régression. La conclusion sera donc évidemment la même que celle obtenue avec la variable indépendante « binaire » ; on ne rejette pas l'hypothèse nulle.

**Tableau 5.6 : Régression linéaire de la variable « Gainrelatif » avec sa variable laguée pour l'équipe des Philadelphia Philies**

`. reg Gainrelatif L1.Gainrelatif`

Source	SS	df	MS	Number of obs	=	1,121
Model	.001977805	1	.001977805	F(1, 1119)	=	0.00
Residual	1308.08742	1,119	1.16897892	Prob > F	=	0.9672
				R-squared	=	0.0000
				Adj R-squared	=	-0.0009
Total	1308.08939	1,120	1.16793696	Root MSE	=	1.0812

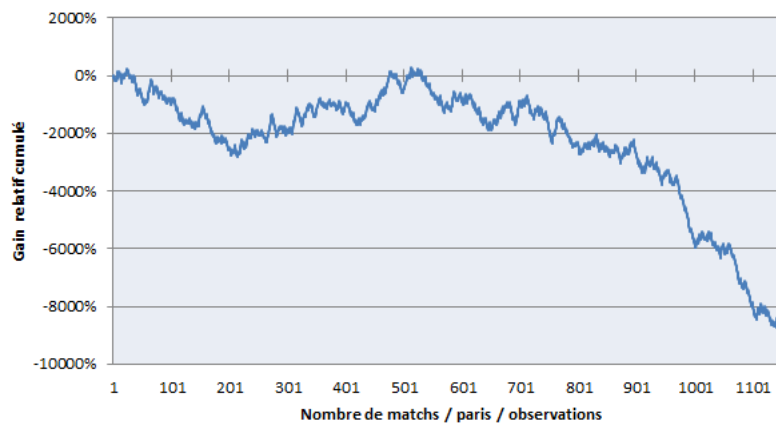
  

Gainrelatif	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Gainrelatif L1.	-.0012301	.0299064	-0.04	0.967	-.0599091 .0574488
_cons	-.0840576	.0323947	-2.59	0.010	-.1476188 -.0204965

### Analyse du cas n°3 : Mise sur la défaite des Boston Red Sox

#### Analyse graphique

Sur la figure 5.5, on fait état d'un constat marquant : Alors que la première moitié des matchs ne s'avère pas être trop mauvaise pour le parieur, avec un gain relatif cumulé proche de 0, la seconde partie des observations démontre une chute impressionnante du gain relatif cumulé atteignant au bout des 6 années un résultat de -9151%. Ceci est dû au fait que les Boston Red Sox se soient renforcés en 2016 et ont dès lors commencé à multiplier les victoires.



*Figure 5.5 : Evolution des gains relatifs cumulés entre 2013 et 2018 (Boston Red Sox)*

Le tableau 5.6 représente le parfait exemple de la distribution des gains relatifs pour une équipe moyenne. En effet, on constate que la majorité des points au delà de 0% se situent aux alentours de 100%, ce qui signifie que la cote des bookmakers était de 2 et donc qu'ils prédisaient une probabilité que les Boston Red Sox perdent de près de 50%. Parfois les gains relatifs passent en dessous de 50% (l'équipe faisait alors office de favori) ou au dessus de 200% (l'équipe était en position de challenger) mais on arrive pas du tout à des valeurs extrêmes comme on a notamment pu observer par ailleurs.

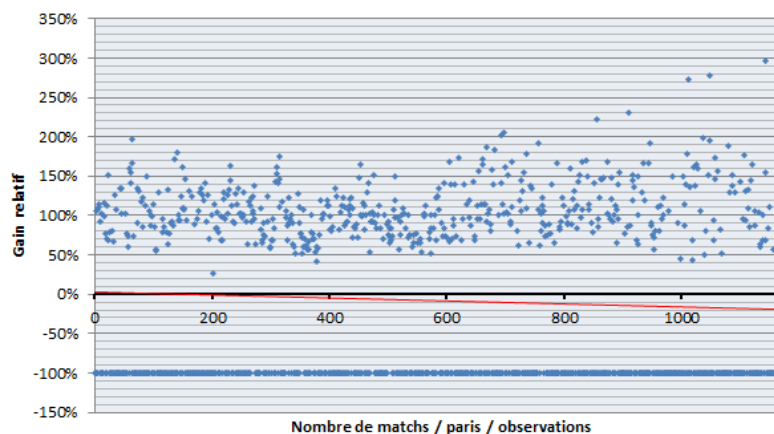


Figure 5.6 : Gain relatif de chaque match entre 2013 et 2018 (Boston Red Sox)

### Run test

Tableau 5.7: Réalisation du Run test pour l'équipe des Boston Red Sox

n1	n2	n	R	E ( R )	Var	StD	$\alpha$	Z	P-value
515	644	1159	579	573,321	282,365	16,804	0,05	0,338	0,735

515 matchs ont été perdus par les Boston Red Sox et ont permis au parieur de gagner de l'argent. Malheureusement, ce n'est manifestement pas assez pour un total de 1159 matchs et on a pu voir sur la figure 5.5 que le résultat final est très négatif au bout du compte pour l'investisseur. La p-value résultant de cette analyse s'élève à 0.735. Il en découle qu'il y a 73.5% de chance d'obtenir cette série d'observations si  $H_0$  est vraie et on va donc déclarer que ces données sont aléatoirement distribuées au seuil de 5%, puisque la p-value est supérieure à ce seuil.

### Régression linéaire

Le coefficient de détermination  $R^2$  étant de l'ordre de 0.0001 dans le tableau 5.8, cela signifie que le modèle est de mauvaise qualité. Le coefficient de -0.010 est très faible et n'est de toute façon pas significatif puisque la p-value qui lui est liée (0.718) est supérieure au seuil de signification  $\alpha$ . Sur base de ces résultats, on ne peut pas rejeter l'hypothèse nulle au seuil de 5%, et on déclare de ce fait que la variable Binaire au temps t-1 n'a aucune influence sur la variable Binaire au temps t.

**Tableau 5.8 : Régression linéaire de la variable « binaire » avec sa variable laguée (Boston Red Sox)**

**. reg Binaire L1.Binaire**

Source	SS	df	MS	Number of obs	=	1,158
Model	.032256554	1	.032256554	F(1, 1156)	=	0.13
Residual	285.93061	1,156	.247344819	Prob > F	=	0.7181
				R-squared	=	0.0001
				Adj R-squared	=	-0.0008
Total	285.962867	1,157	.247158917	Root MSE	=	.49734

Binaire	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Binaire L1.	-.0106207	.0294101	-0.36	0.718	-.0683239	.0470824
_cons	.4494557	.0196131	22.92	0.000	.4109744	.4879369

Le tableau ci-dessous montre, quant à lui, un R-squared totalement nul, un coefficient très proche de 0 ainsi qu'une p-value associée très élevée, ce qui nous permettra de conclure sans trop d'hésitations que ce modèle est nul, qu'il n'est pas possible, dans de telles conditions, de rejeter  $H_0$ , et ainsi d'affirmer que  $GR_{t-1}$  n'a aucun pouvoir explicatif sur  $GR_t$ .

**Tableau 5.9 : Régression linéaire de la variable « Gainrelatif » avec sa variable laguée pour l'équipe des Boston Red Sox**

**. reg Gainrelatif L1.Gainrelatif**

Source	SS	df	MS	Number of obs	=	1,158
Model	.014675297	1	.014675297	F(1, 1156)	=	0.01
Residual	1323.36989	1,156	1.14478364	Prob > F	=	0.9099
				R-squared	=	0.0000
				Adj R-squared	=	-0.0009
Total	1323.38457	1,157	1.14380689	Root MSE	=	1.0699

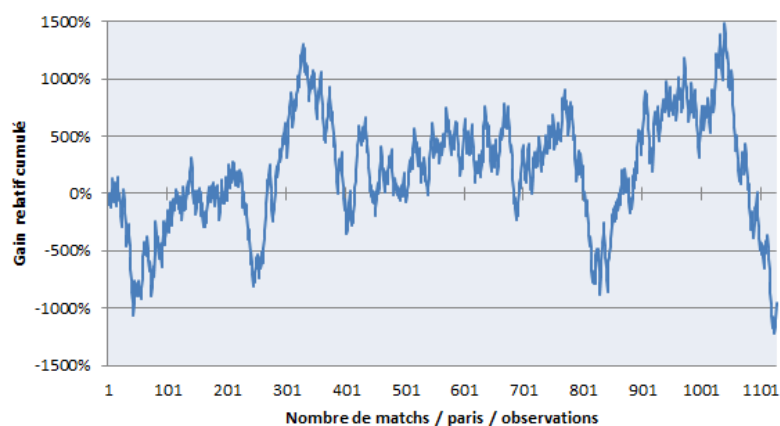
Gainrelatif	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Gainrelatif L1.	-.00333	.0294116	-0.11	0.910	-.0610361	.0543761
_cons	-.0836195	.0315372	-2.65	0.008	-.1454961	-.0217429

## Analyse du cas n°4 : Mise sur la défaite des Colorado Rockies

### Analyse graphique

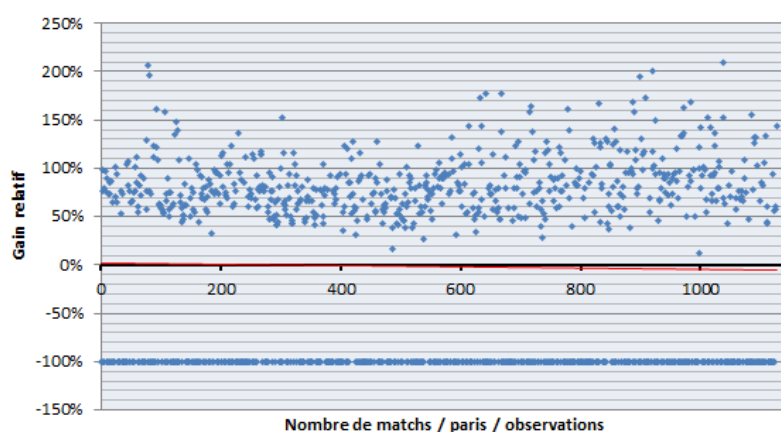
La figure 5.7 met en évidence un scénario particulier que nous n'avons pas observé précédemment. En effet, dans le cadre de cette analyse, on peut apercevoir que le gain relatif cumulé ne cesse de fluctuer entre valeur positive et négative, ce qui était rarement le cas pour les autres équipes. De plus, il est important de signaler que l'échelle de ce graphique est

beaucoup plus concentrée puisque le gain relatif cumulé ne fluctue qu'entre 1500% et -1500%. Au final, malgré une longue période de rentabilité pour le parieur, la dernière centaine de matchs aura une influence négative sur le gain relatif cumulé à tel point qu'il sera de près de -1000% à l'échéance des 6 années.



*Figure 5.7 : Evolution des gains relatifs cumulés entre 2013 et 2018 (Colorado Rockies)*

Ce qui était vrai pour la figure 5.6 est d'autant plus vrai pour la figure 5.8. Encore une fois on assiste à un nuage de points représentant typiquement les gains relatifs d'une équipe moyenne car ils se situent principalement autour de 100%, soit une cote de 2 pour le bookmaker. Le nombre important de points démontre que les défaites ont été nombreuses au cours des 6 années et aucun gain relatif n'est supérieur à 250%, ce qui prouve que l'équipe n'est jamais considérée comme étant la grande favorite du match.



*Figure 5.8 : Gain relatif de chaque match entre 2013 et 2018 (Colorado Rockies)*

## Analyse du run test

**Tableau 5.10 : Réalisation du Run test pour l'équipe des Colorado Rockies**

n1	n2	n	R	E ( R )	Var	Std	$\alpha$	Z	P-value
609	518	1127	544	560,826	277,838	16,668	0,05	1,009	0,313

Sur un total de 1127 observations, la réalisation du run test de cette série de données converge vers une p-value de 0.313. Puisqu'elle est supérieure au seuil  $\alpha$ , on décide de ne pas rejeter l'hypothèse nulle stipulant que les variables sont aléatoirement distribuées et on convient de ce fait que la séquence observée suit une distribution aléatoire.

## Régression linéaire

En ce qui concerne le tableau 5.11, la somme expliquée des carrés (0.238) étant dérisoire par rapport à la somme résiduelle des carrés (279,382), cela implique que le R-squared du modèle soit très faible. Cela désigne que la qualité du modèle n'est pas bonne.

Par ailleurs, lorsque l'on s'attarde sur la variable  $CB_{t-1}$ , représentée sur ce tableau par la variable Binaire L1, on se rend compte que le coefficient qui y est associé est de 0.029. Autrement dit, l'influence de la variable « code binaire » à l'instant t-1 sur la variable code binaire à l'instant t est très faible voire nulle. La p-value de 0.328 nous indique que le résultat n'est pas significatif à un seuil de 5% et qu'il ne faut par conséquent pas rejeter l'hypothèse nulle. On peut alors en conclure que  $CB_{t-1}$  n'a aucun pouvoir explicatif sur  $CB_t$ .

**Tableau 5.11: Régression linéaire de la variable « binaire » avec sa variable laguée pour l'équipe des Colorado Rockies**

**. reg Binaire 1.Binaire**

Source	SS	df	MS	Number of obs	=	1,126
Model	.238154775	1	.238154775	F(1, 1124)	=	0.96
Residual	279.382627	1,124	.248561056	Prob > F	=	0.3279
				R-squared	=	0.0009
				Adj R-squared	=	-0.0000
Total	279.620782	1,125	.248551806	Root MSE	=	.49856

Binaire	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Binaire L1.	.0291798	.0298105	0.98	0.328	-.0293107 .0876703
_cons	.5250965	.0219054	23.97	0.000	.4821164 .5680766



Avec un coefficient  $\beta$  de l'ordre de 0.029, une p-valeur non-significative de 0.322 et une variable explicative  $R^2$  nulle, la régression présente ci-dessous ne déroge pas à la règle et vient s'inscrire dans la lignée de tout ce que l'on a pu constater auparavant ; On ne rejette pas l'hypothèse nulle et on considère que  $R_{t-1}$  n'a aucun pouvoir explicatif sur  $R_t$ .

**Tableau 5.12 : Régression linéaire de la variable « Gainrelatif » avec sa variable laguée pour l'équipe des Colorado Rockies**

**. reg Gainrelatif L1.Gainrelatif**

Source	SS	df	MS	Number of obs	=	1,126
Model	.908504771	1	.908504771	F(1, 1124)	=	0.98
Residual	1040.99786	1,124	.926154678	Prob > F	=	0.3222
				R-squared	=	0.0009
				Adj R-squared	=	-0.0000
Total	1041.90636	1,125	.926138989	Root MSE	=	.96237

Gainrelatif	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Gainrelatif L1.	.0295454	.029831	0.99	0.322	-.0289854	.0880762
_cons	-.0150616	.0286845	-0.53	0.600	-.0713427	.0412195



## Chapitre 6 - Influence des news sur le marché des paris sportifs

Ce chapitre nous donne l'occasion d'entrevoir l'influence de l'information et des « news » sur le marché des paris sportifs. En effet, la question de savoir si un agent «informé» a une capacité supérieure à traiter des informations publiques est centrale et vient s'imbriquer dans notre analyse de l'efficacité de ce marché. Comme nous l'avons vu au cours du chapitre 2, nous cherchons ici à savoir si le marché des paris sportifs est efficace au sens « semi-fort ».

Que ce soit pour un parieur ou pour un bookmaker, l'information s'avère être une véritable source de valeur puisque le fait de disposer d'une quantité supérieure d'informations ou de meilleure qualité permet généralement à l'agent d'avoir un avantage non-négligeable par rapport à l'agent non informé. Par conséquent, nous tentons dans ce chapitre de savoir s'il existe une asymétrie d'information (publique) entre parieurs et bookmakers, au sein du marché des paris sportifs et si, tel est le cas, s'il est possible pour un parieur d'en tirer profit.

Le parallèle avec le marché financier est relativement aisé puisque, autant dans le monde de la bourse que des paris sportifs, il s'agit un peu de la course à l'information. L'information, bonne ou mauvaise, doit être perçue le plus vite possible et puis immédiatement répercutée sur la décision d'investissement la plus pertinente. Dans le cas du scandale VW par exemple, certains ont moins perdu que d'autres, uniquement grâce à leur capacité à capter l'information et d'agir en fonction de celle-ci. Pour les paris sportifs c'est un peu pareil. Si un consultant très expérimenté venait à prédire la victoire d'un tel coureur cycliste, la cotation ne tarderait pas à diminuer car la probabilité associée au fait qu'il gagne augmente (ou du moins la perception qu'on a qu'il gagne augmente, puisque, comme nous l'avons vu auparavant, « les bookmakers ne proposent pas des côtes reflétant la force réelle des opposants, mais plutôt des côtes reflétant l'opinion des parieurs quant aux forces de ces opposants »).

Alors que l'information est ainsi source d'une volatilité changeante au sein des marchés financiers, il en va de même pour le marché des paris sportifs. Dans le chapitre 1, on a vu que, de manière théorique la formule 1.1 s'appliquait. Pour rappel on avait :

$$C_i = 1/p$$

Ensuite, on a constaté que cela ne s'avérerait pas être tout à fait juste puisque de nombreux ajustements devaient être effectués après avoir trouvé une certaine cote initiale théorique. En effet, on avait vu que cote devait être modifiée en fonction des news, de la cotation des

bookmakers et des mises des joueurs. Pour le bookmaker, l'ensemble de ces 3 facteurs à prendre en considération n'est rien autre que de l'information.

Par conséquent, ceci n'est autre que de dire que :

$$C_i = 1/p + \sum_{i=1}^q N_i \quad (6.1)$$

où  $N_i$  représente une nouvelle information quelconque ayant un impact direct sur la cote ( $C(i)$ ) et où  $\sum_{i=1}^q N_i$  représente la somme de  $p$  news ayant un impact direct sur la cote ( $C(i)$ ).

Or cette formule n'est pas nouvelle, et fait référence à ce qu'ont découverts CLARK (1973) EPPS (1976) et TAUCHEN et PITTS (1983) pour les marchés financiers en testant la relation entre le volume et la volatilité. Soit :

$$r_t = \mu + \sum_{i=1}^{N_t} \varepsilon_{i,t} \quad (6.2)$$

Le rendement ( $r_t$ ) d'un jour donné n'est rien d'autre que l'agrégation d'une constante ( $\mu$ ) à une somme de news ( $\sum_{i=1}^{N_t} \varepsilon_{i,t}$  qui est indépendante et identiquement distribuées avec une moyenne 0 et une variance égale à  $\delta^2$ ). On observe qu'il en va de même pour le marché des paris sportifs. La constante  $\mu$  est représentée par l'inverse de la probabilité d'occurrence de l'évènement ( $1/p$ ), qui était calculée, si l'on se souvient de ce qui a été dit dans le chapitre 1, grâce aux algorithmes et aux banques de données des bookmakers. Cette constante ( $\mu$ ) fait donc référence à notre « cote initiale théorique ». A cela, vient s'ajouter un ajustement dû aux nouvelles.

La théorie de Clark reposant sur l'idée que toute arrivée d'information sur un marché financier engendre une variation instantanée des cours des titres et du nombre d'échanges, une augmentation du flux d'information ferait donc en sorte d'intensifier le volume de l'actif ainsi que les mouvements des prix de cet actif. Ceci vaut également pour le marché des paris sportifs puisque, si de nombreuses informations publiques sont perçues, cela va avoir pour effet d'augmenter le nombre de mises puisque les parieurs vont commencer à spéculer davantage sur un scénario bien spécifique résultant de l'information captée. Les mises, vont, elles, également s'ajuster beaucoup plus souvent aux nouvelles informations publiques puisque celle-ci sont intégrées par le bookmaker dans leurs cotations, comme nous allons le voir ci-après. Ainsi nous venons de prouver qu'un parallèle existe entre les marchés financiers et le marché des paris sportifs en ce qui concerne le rôle joué par les « news » sur la volatilité et le volume des cotations/actifs.

## Influence des informations publiques sur la cote « live »

L'analyse ci-dessous consiste à prouver que les agents intègrent effectivement les données publiques pour fixer les cotations des événements. Alors que nous nous sommes concentrés sur la cotation en pré-match lors de notre analyse de l'efficacité faible, nous allons, désormais, dans le cadre de cette analyse ci, considérer les cotations dites « live ». Les cotations live permettent de placer en permanence une mise sur un événement qui est en train de se dérouler. Ainsi, une fois un événement sportif commencé, les cotes de pré-match en lien avec cet événement se terminent au profit des cotations live qui prennent place sur les plateformes en ligne. Alors que les cotes pré-match varient à peu de reprises, les cotations en live ont, elles, la particularité de fluctuer de manière incessante de telle sorte qu'il n'est pas rare de voir les cotes live d'un événement sportifs changer plus de 300 fois pendant qu'il a lieu.

Les données utilisées pour la réalisation de cette analyse proviennent de tableaux de cotation minute par minute fournies par l'entreprise « Ardent Groupe » sur le match de football opposant Schalke 04 à Manchester City. Dans ces bases de données, on retrouve l'évolution des cotes minute par minute pour chaque scénario possible (la victoire de Manchester City, le match nul et la victoire de Schalke 04)

Afin d'illustrer mes propos, voici, dans le tableau 6.1, une ébauche des données qui ont été traitées (les données pré-match sur le dessus du tableau 6.1 n'ont pas été prises en compte)

**Tableau 6.1 : Ebauche des données collectées avec l'évolution des cotes minute par minute pour le scénario « Victoire de Manchester City » lors de la confrontation entre Schalke 04 et Manchester City le 20 février 2019.**

Equipe	Cote	Statut	Date	Heure	Live / Prematch
Manchester City	1,31	Open	20/02/2019	18:54:07	Prematch
Manchester City	1,30	Open	20/02/2019	19:02:08	Prematch
Manchester City	1,31	Open	20/02/2019	19:02:21	Prematch
Manchester City	1,30	Open	20/02/2019	19:04:17	Prematch
Manchester City	1,31	Open	20/02/2019	19:10:26	Prematch
Manchester City	1,29	Open	20/02/2019	19:49:15	Prematch
Manchester City	1,33	Open	20/02/2019	19:49:25	Prematch
Manchester City	1,33	Open	20/02/2019	19:56:19	Prematch
Manchester City	1,33	Open	20/02/2019	20:01:55	Live
Manchester City	1,29	Open	20/02/2019	20:02:25	Live
Manchester City	1,29	Open	20/02/2019	20:02:30	Live
Manchester City	1,29	Open	20/02/2019	20:02:34	Live
Manchester City	1,33	Open	20/02/2019	20:03:20	Live
Manchester City	1,33	Open	20/02/2019	20:04:36	Live
Manchester City	1,33	Open	20/02/2019	20:06:31	Live
Manchester City	1,33	Open	20/02/2019	20:07:08	Live
Manchester City	1,29	Open	20/02/2019	20:08:04	Live
Manchester City	1,30	Open	20/02/2019	20:08:59	Live
Manchester City	1,33	Open	20/02/2019	20:09:15	Live
Manchester City	1,29	Open	20/02/2019	20:11:54	Live
Manchester City	1,29	Open	20/02/2019	20:12:09	Live

Par ailleurs, le tableau 6.2 reprend, quant à lui, les principaux faits majeurs du match en question, recueillis sur le site [lequipe.fr](http://lequipe.fr)<sup>12</sup>, qui sont susceptibles d'influencer de manière assez conséquente la cotation. Les plus petits faits de match (comme les coup-francs, les cartons jaunes, les corners ou les phases dangereuses) n'ont pas été repris car il est plus difficile de voir s'il y a un lien de cause à effet entre l'action et la cote.

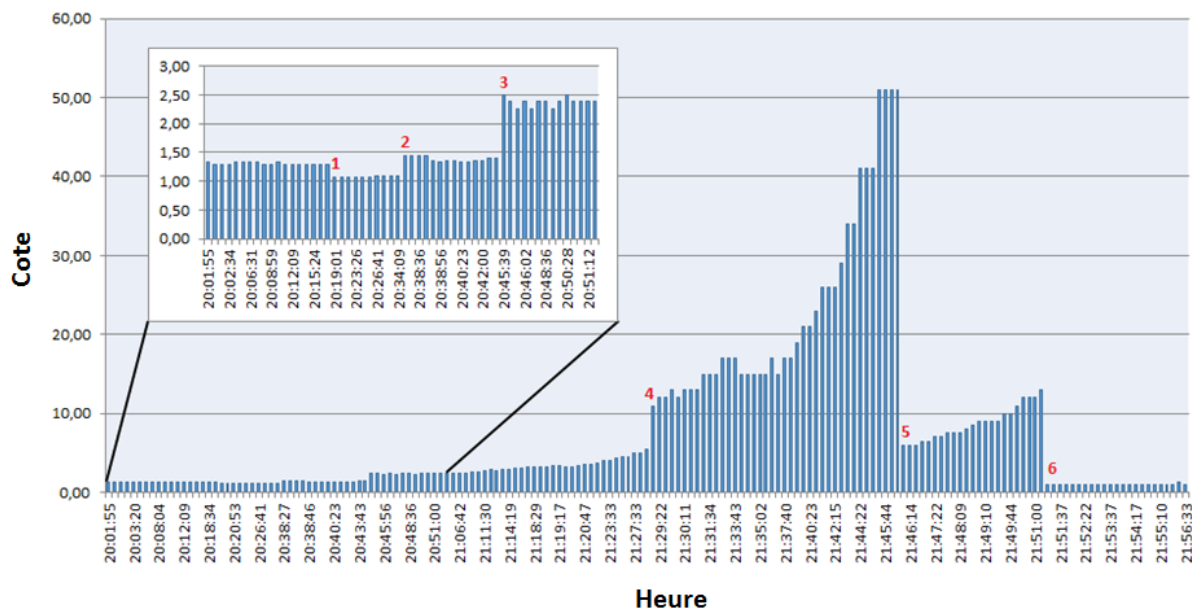
**Tableau 6.2 : Faits de match majeurs lors de la confrontation entre Schalke 04 et Manchester City le 20 février 2019**

N°	Fait de match	Minute de jeu	Score
1	But pour Manchester City	18'	0-1
2	But pour FC Schalke 04	38'	1-1
3	But pour FC Schalke 04	45'	2-1
4	Carton rouge pour Manchester City	68'	2-1
5	But pour Manchester City	85'	2-2
6	But pour Manchester City	90'	2-3

Grâce aux données fournies par la société Groupe Ardent pour le match entre Schalke 04 et Manchester City (présentes en ébauche dans le tableau 6.1) il m'a été possible de reproduire graphiquement l'évolution des cotes minute par minute pour chacun des 3 scénarios possibles.

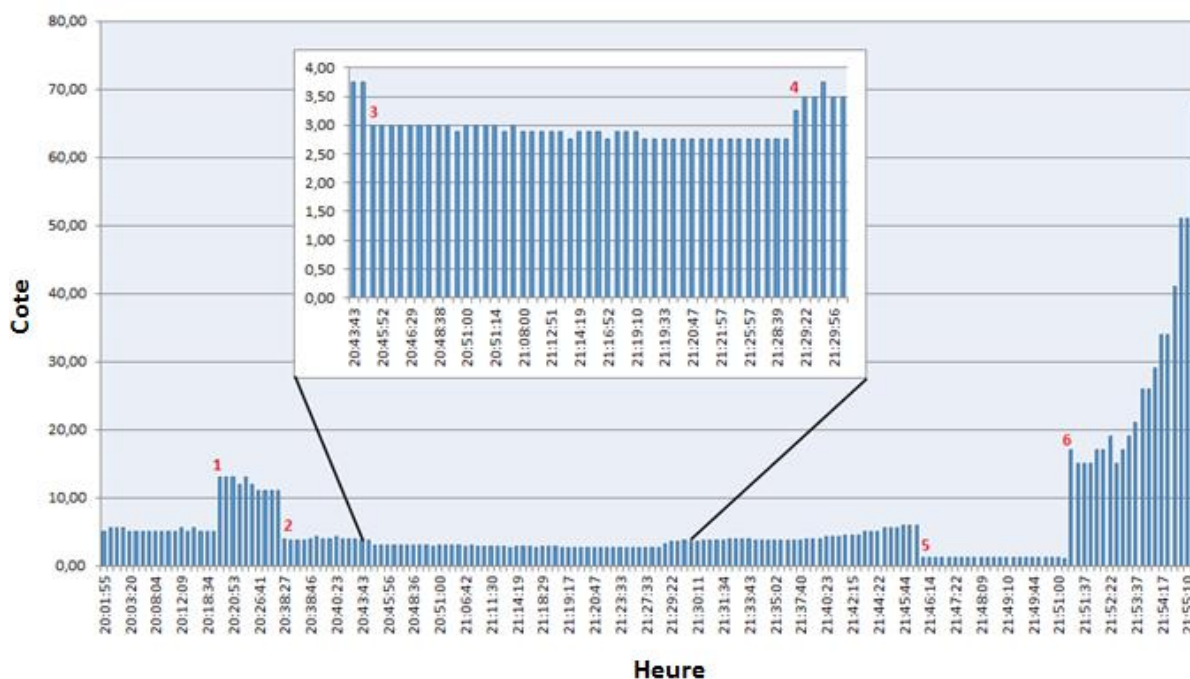
La figure 6.1 représente l'évolution des cotes du match pour le scénario « Victoire de Manchester City ». En rouge, sont représentés les 6 faits de match mentionnés dans le tableau 6.2 et il est donc possible, via la figure 1, d'observer l'incidence de ces faits de match sur les cotations. Les faits de match (1), (5) et (6) du tableau 6.2 sont bénéfiques pour Manchester City et augmentent leur probabilité de gagner le match. Ainsi, le bookmaker diminue la cotation associée au scénario « Victoire de Manchester City » lorsque ces 3 faits ont lieu. On voit donc par exemple que la cotation, en 5 minutes, passe de 51 à 1.04, suite à l'inscription de 2 buts de la part de Manchester de City en toute fin de rencontre. Avec un score en défaveur de 2 buts à 1 pour Schalke 04, la probabilité de voir gagner Manchester City avant le but égalisateur était de moins de 2% ! Quelques minutes plus tard, la probabilité de voir Manchester City gagner approche les 100% puisqu'il ne reste que très peu de temps à jouer (2-3 minutes tout au plus).

<sup>12</sup> <https://www.lequipe.fr/Football/match-direct/ligue-des-champions/2018-2019/schalke-manchester-city-live/447939>



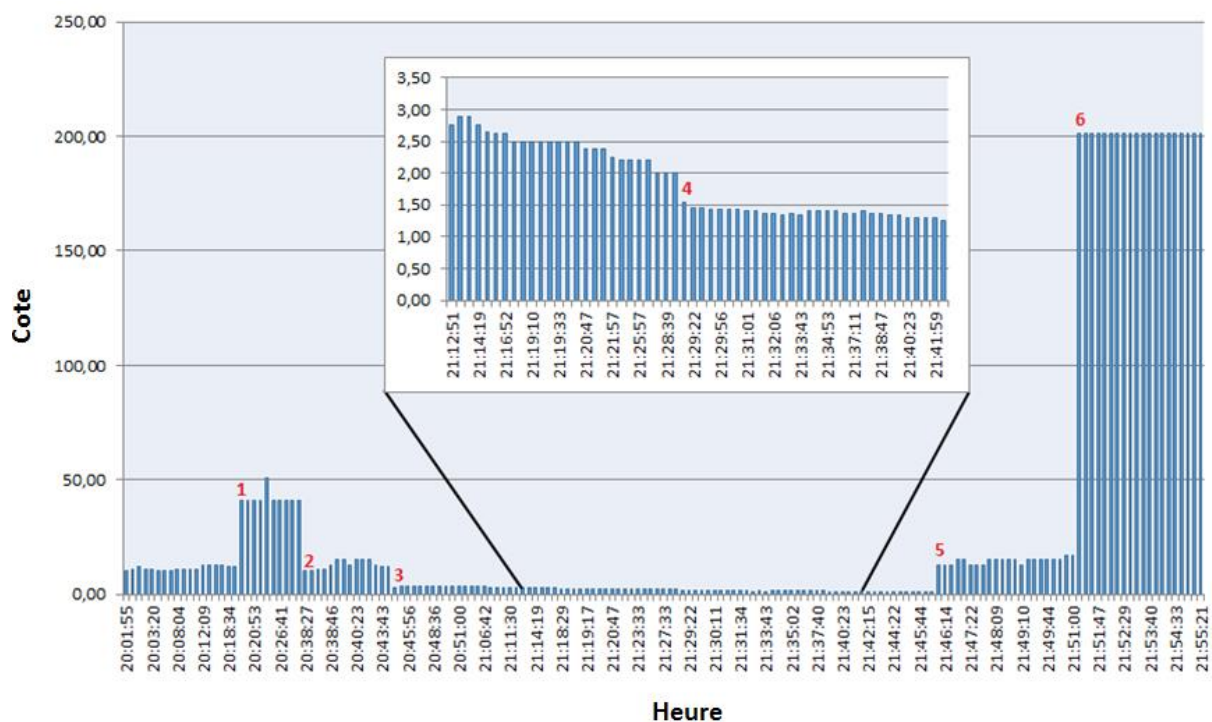
*Figure 6.1 : Evolution des cotes minute par minute du scénario : « Victoire de Manchester City » lors de la confrontation entre Schalke 04 et Manchester City le 20 février 2019*

La seconde figure nous offre une perspective de l'évolution de la cotation pour le scénario « Match nul ». Puisque nous sommes ici dans le cadre d'un match où il n'y a jamais plus de deux buts d'écart, les cotes associées au match nul de ce match ne sont jamais très élevées. Le seul moment où la cote prend de l'ampleur, c'est au moment même où Manchester City inscrit l'ultime but de la rencontre à la 90<sup>ème</sup> minute, alors qu'il ne reste que 5 minutes de temps additionnels. On aperçoit ainsi que la cote passe de 1.11 à 17 (6) suite à ce goal. Cela s'explique par le fait qu'à la 89<sup>ème</sup> minute de jeu, le score était toujours de 2-2 et la probabilité de voir le match se finir sur ce résultat était très élevée (90%), d'où la cote de 1.11. Une fois le troisième but de Manchester City inscrit, le résultat n'est plus un match nul, et vu l'état avancé du match, la cote s'envole minute après minute pour atteindre, une minute avant le coup de sifflet final, une cote de l'ordre de 67.



*Figure 6.2: Evolution des cotes minute par minute du scénario « Match nul » lors de la confrontation entre Schalke 04 et Manchester City le 20 février 2019*

Le troisième et dernier scénario est représenté sur la figure 6.3. Elle représente l'évolution des cotes minute par minute pour le scénario relatif à la victoire de l'équipe Schalke 04. La fluctuation des cotes pour ce scénario est impressionnante. La cote de début de match est de 10 pour ce scénario, soit 10% de chance que Schalke 04 remporte la rencontre selon les bookmakers. Le but de Manchester City (1) à la 18<sup>ème</sup> minute de jeu vient tasser la probabilité à seulement 2.4% puisque le bookmaker vient ajuster sa cote à 41. Ensuite les deux buts (2 & 3), coup sur coup, de Schalke 04 viennent au contraire redescendre progressivement la cote à une cote de 3. Le carton rouge reçu par un joueur de Manchester City (4) ne viendra modifier que très légèrement la cote pour ce scénario. Lors de l'égalisation à « 2 partout », la cote grimpe à 13 alors qu'elle monte carrement à une cote de 200 lors du troisième but de Manchester City (soit une probabilité de voir Schalke 04 remporter le match de seulement 0.5% !). En effet, il faudrait que Schalke 04 puisse inscrire 2 buts en l'espace de moins de 5 minutes pour que ce scénario soit réalisé, ce qui est très peu probable au vue de la configuration du match.



*Figure 6.3 : Evolution des cotes minute par minute du scénario « Victoire de Schalke 04 » lors de la confrontation entre Schalke 04 et Manchester City le 20 février 2019*

On aperçoit donc clairement que l'ajustement est effectué en live afin d'ajuster les cotes aux probabilités reflétant les chances exactes de chaque scénario de se produire. Chacun des 6 faits majeurs implique un ajustement plus ou moins grand de la cote. On peut donc en conclure que même en live, les bookmakers intègrent l'information publique dans leurs cotations.

### **Vitesse d'ajustement de la cote en live**

Alors que nous venons de démontrer qu'un ajustement de la cote est bel et bien réalisé de la part des bookmakers lors de rencontres en live, tentons désormais d'observer à quelle vitesse elle est ajustée.

En effet, cette vitesse d'ajustement est capitale pour les bookmakers car, tant bien même qu'ils ajustent leurs cotes en fonction des données publiques, si le parieur est plus rapide que le bookmaker pour appréhender ces informations publiques, le marché n'est pas efficient au sens semi-fort, puisque le parieur pourrait en tirer profit au détriment des bookmakers.

Partons une nouvelle fois des données fournies par le groupe Ardent pour mener à bien cette analyse. Le tableau 6.3 reprend l'évolution de la cotation avant, pendant et après l'obtention



du carton rouge pour Manchester City (fait majeur (4) dans le tableau 6.2 et numéro (4) sur la figure 6.1). La cote de Manchester City s'élève à 5.5 à 21 h 28 min 57 s. Entre ce moment et 21 h 29 min 01 s, un carton rouge est attribué à un des joueurs de Manchester City. Cela signifie qu'il a fallu, au maximum, 4 secondes au bookmaker pour suspendre provisoirement les paris en ligne. La suspension dure dans ce cas-ci 15 secondes<sup>13</sup>. 15 secondes durant lesquels il est tout à fait impossible pour le parieur de placer une mise.

Cette suspension est nécessaire pour deux raisons principales. La première, on vient de l'évoquer, consiste à empêcher le parieur de miser sur une cotation qui n'est plus d'actualité puisqu'une nouvelle information publique est venue modifier les probabilités d'occurrences des scénarios de cet évènement sportif. Le deuxième facteur justifiant cette suspension provisoire de paris, est justement que le bookmaker doit trouver la cote qui reflète au mieux la nouvelle probabilité d'occurrence. Modifier rapidement l'ancienne cote pour empêcher le parieur de miser, tout en proposant une cote qui n'est pas du tout en adéquation avec les probabilités réelles d'occurrence n'aurait en effet pas de sens.

**Tableau 6.3 : Evolution des cotations du scenario « Victoire de Manchester City » pour le fait de match « Carton rouge pour Manchester City » (4) du tableau 6.2.**

Equipe	Cote	Statut	Date	Heure
Manchester City	5,5	Open	20-02-19	21.28.57
Manchester City	1.00	Suspended	20-02-19	21.29.01
Manchester City	11	Open	20-02-19	21.29.16

En prenant un exemple similaire avec la cotation de Schalke 04 lors de l'inscription du deuxième but de Manchester City (fait majeur n°5), on aperçoit, dans le tableau 6.4 que l'ajustement se fait en un éclair. Il n'a fallu que 6 secondes au bookmaker pour modifier la cote, et ainsi prendre en compte l'information publique dans la cotation.

**Tableau 6.4 : Evolution des cotations du scenario « Victoire de Schalke 04 » pour le fait de match « But pour Manchester City » (5) du tableau 6.2.**

Equipe	Cote	Statut	Date	Heure
FC Schalke 04	1,18	Open	20-02-19	21.45.49
FC Schalke 04	1.00	Suspended	20-02-19	21.45.51
FC Schalke 04	13	Open	20-02-19	21.45.55

On vient de le voir, le bookmaker est ultra réactif quant à l'ajustement de la cotation lorsque de nouvelles informations publiques émanent lors d'évènements sportifs « live ». En

<sup>13</sup> La durée (relativement longue) de cette suspension des paris peut s'expliquer par le fait qu'il s'agit ici d'un carton rouge et qu'il est possible que la vidéo-assistance (VAR) vienne contredire la décision de l'arbitre en annulant cette carte rouge. Visiblement, après ces 15 secondes, le bookmaker était persuadé qu'il n'y aurait pas de retour en arrière, pouvant ainsi valider sa nouvelle cote.



suspendant les mises pendant quelques secondes à des moments cruciaux, un goal, un penalty, un carton rouge mais aussi plus récemment l'utilisation de la vidéo (VAR), le bookmaker se prévaut également de toute tentative de stratégies provenant des parieurs visant à être plus réactif que le bookmaker.

Par ailleurs, les bookmakers mettent en place des stratégies qui les assurent qu'il soit impossible de tirer profit d'un éventuel laxisme de l'ajustement de la cotation. Par exemple, les sites des bookmakers sont faits de telle sorte qu'il n'est pas possible de placer une mise en 2-3 secondes. Il faut plusieurs secondes pour choisir sa mise, la valider et la payer, ce qui offre encore au bookmaker un délai supplémentaire.

## Conclusion

### Conclusion sur l'efficience faible

Les analyses menées sur les équipes de basketball et de baseball ont été concluantes et nous ont permis de tester l'efficience faible du marché des paris sportifs pour ces 2 sports. Pour rappel, l'objectif de cette recherche était de prouver que l'information contenue dans la série historique des prix est reflétée dans les cotes de telle manière qu'un agent ne peut pas tirer profit d'informations passées pour prédire les rendements futurs.

D'un point de vue théorique, nous savions déjà que, puisque les bookmakers admettent consulter des bases de données historiques pour forger leur cote initiale théorique, les cotes reflètent déjà toute l'information contenue dans la série historique des prix. D'un point de vue pratique, les différents tests ont mené à plusieurs conclusions.

Les analyses graphiques ont reflété un rendement cumulé négatif pour l'ensemble des équipes analysées. Il semble dès lors qu'opter pour une même stratégie de mise à long terme, n'est pas profitable pour le parieur. Nous avons vu que même dans des conditions très favorables (60% de paris gagnants), le gain relatif cumulé s'avère être toujours négatif. De manière générale, on n'observe aucune différence entre les différents scénarios (mise sur la victoire ou sur la défaite de l'équipe) ou les différents sports (baseball et basketball). Il est vrai que les pertes s'avèrent être plus élevées pour le baseball que pour le basketball, mais puisque le nombre de matchs analysés passe du simple au double pour le baseball, on ne se permettra pas de conclure quoi que ce soit à ce propos.

La réalisation de runs tests a permis de constater que les variables étaient effectivement distribuées de manière aléatoire, et, de ce, fait, qu'il n'existe pas de tendances pouvant être exploitées par le parieur pour adopter des stratégies qui lui permettrait de réaliser des profits à long terme.

L'étude des régressions linéaires est, quant à elle, venue confirmer ces propos puisqu'on a pu observer, dans un premier temps, qu'il n'y avait pas de lien entre le résultat qualitatif (perte ou gain d'argent) antérieur et le résultat qualitatif présent/futur et, dans un second temps, que le gain relatif antérieur, c'est-à-dire le rendement antérieur, n'expliquait nullement le rendement actuel/futur.

De par ces analyses, on peut ainsi en conclure que, sur base des données analysées, le baseball et le basketball sont deux sports pour lesquels le marché des paris sportifs est efficient au sens faible du terme. Ceci étant dû au fait que l'analyse des rendements passés ne permet pas de prévoir les rendements futurs puisque les probabilités implicites reflètent déjà toutes les informations antérieures pertinentes.

### **Conclusion sur l'efficience semi-forte**

Lors de l'analyse de l'efficience semi-forte, l'objectif était de démontrer que le bookmaker prenait en considération les informations publiques dans sa cotation, et ce, de manière instantanée. L'étude du match de football nous a permis de voir qu'un ajustement de la cotation était bel et bien réalisé et que celui-ci était pertinent par rapport aux informations publiques perçues. Nous avons également observé que la vitesse de l'ajustement était tel, qu'il n'était pas possible pour un parieur de profiter de ces informations publiques pour réaliser un profit. Ceci étant également renforcé par les mesures de blocage ou de suspension de la cote prises par le bookmaker. Le marché des paris sportifs s'avère ainsi être efficient au sens semi-fort du terme puisque la cote assimile déjà l'ensemble des informations disponibles au grand public.

### **Conclusion générale et perspectives**

A travers ce mémoire, on a pu constater que le marché des paris sportifs s'avère être efficient puisqu'il est impossible de mettre en place des stratégies qui permettent d'obtenir des rendements positifs à long terme. La question de savoir s'il est possible de battre les bookmakers a ici pris une tournure plutôt négative. Néanmoins, ce mémoire ne couvre pas l'analyse de l'efficience forte, et il serait intéressant de se pencher sur la question car si le marché des paris sportifs est inefficent, cela ne peut certainement provenir que des asymétries d'informations privées. Seules ces dernières sont susceptibles de mettre à mal les bookmakers, mais la démonstration de cette hypothèse suppose des recherches poussées, des enquêtes de terrain ou des témoignages de la part de personnes ayant disposé d'informations privées et en ayant tiré profit. Frôlant

Suite à la réalisation de ce mémoire, d'autres recherches pourraient par contre être menées afin de compléter les conclusions qui débouchent de ce travail. Par exemple l'étude des paramètres influençant la cotation d'un événement sportif pourrait être réalisée. Ainsi, cela permettrait par exemple de savoir quel poids possède la partie « mécanisme de détermination

d'une cote initiale théorique » par rapport au « mécanisme d'ajustement de la cote ». Est-ce que l'algorithme est en mesure d'expliquer un haut pourcentage de la cotation finale ? Ou cette cotation finale est-elle majoritairement déterminée par les bookmakers concurrents ou l'apparition de news ?

Par ailleurs, l'analyse de l'ajustement de la cote s'est concentrée sur la partie « live ». Cependant, un travail entier mériterait de s'intéresser sur l'ajustement des cotations en situation de pré-match. Dans quelle mesure et à quelle vitesse le bookmaker prend-il en considération une information publique n'ayant pas lieu pendant le match (limogeage d'un entraîneur, blessure et absence d'un joueur clé, situation familiale compliquée pour un joueur, etc...) et le répercute-t-il sur la cotation de pré-match ?

Au vue de l'émergence des paris sportifs au cours des dernières années, il est de toute façon fort à parier que le nombre de revues scientifiques ne vont cesser de prendre de l'ampleur dans les années qui viennent afin de venir confirmer ou infirmer les propos tenus dans la revue littéraire propre à ce marché.

## Bibliographie

- Alberola, J. M., Garcia-Fornes, A., & Espinosa, A. (2010, September). Price prediction in sports betting markets. In *German Conference on Multiagent System Technologies* (pp. 197-208). Springer, Berlin, Heidelberg.
- Bachelier, L. (1900). Théorie de la spéculation. In *Annales scientifiques de l'École normale supérieure* (Vol. 17, pp. 21-86).
- Barraud, C. (2012). *L'Effcience informationnelle du marché des paris sportifs: un parallèle avec les marchés boursiers*(Doctoral dissertation, Université Paris Dauphine-Paris IX).
- Baxter, L. G. (2011). Betting big: value, caution and accountability in an era of large banks and complex finance. *Rev. Banking & Fin. L.*, 31, 765.
- Berg, J. E., Nelson, F. D., & Rietz, T. A. (2008). Prediction market accuracy in the long run. *International Journal of Forecasting*, 24(2), 285-300.
- Breuer, W., Hauten, G., & Kreuz, C. (2009). Financial instruments with sports betting components: Marketing gimmick or a domain for behavioral finance?. *Journal of Banking & Finance*, 33(12), 2241-2252.
- Brown, A. (2012). Evidence of in-play insider trading on a UK betting exchange. *Applied Economics*, 44(9), 1169-1175.
- Brown, A., & Yang, F. (2017). The role of speculative trade in market efficiency: Evidence from a betting exchange. *Review of Finance*, 21(2), 583-603.
- Cantinotti, M., Ladouceur, R., & Jacques, C. (2004). Sports betting: Can gamblers beat randomness?. *Psychology of addictive behaviors*, 18(2), 143.
- Carcassonne, G. (2011). De quelques mystères du jeu. *Pouvoirs*, (4), 15-23.
- Champagne, L., Gerville-Réache, L., & Tiarks, S. (2014, June). Construction de cote: l'exemple du pari à handicap au Top 14. In *46ème Journées de statistique* (p. 6p)
- Chiappori, P., Jullien, B., Salanie, B., & Salanie, F. (2002). Asymmetric Information in Insurance: some testable implications. Working paper.
- Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: journal of the Econometric Society*, 135-155.
- Costes, J. M., Eroukmanoff, V., & Tovar, M. L. (2014). Les joueurs de paris sportifs et hippiques en ligne. *Les notes de l'Observatoire des jeux*, (4).
- Croxson, K., & James Reade, J. (2013). Information and efficiency: Goal arrival in soccer betting. *The Economic Journal*, 124(575), 62-91.

- Deschamps, B., & Gergaud, O. (2012). Efficiency in betting markets: evidence from English football. *The Journal of Prediction Markets*, 1(1), 61-73.
- Epps, T. W., & Epps, M. L. (1976). The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis. *Econometrica: Journal of the Econometric Society*, 305-321.
- Fama, E. F. (1991). Efficient capital markets: II. *The journal of finance*, 46(5), 1575-1617.
- Figlewski, S. (1979). Subjective information and market efficiency in a betting market. *Journal of political economy*, 87(1), 75-88.
- Fingleton, J., & Waldron, P. (1997). *Optimal determination of bookmakers' betting odds: theory and tests*. Centre for Economic Policy Research.
- Fortis, E. (2009). L'addiction aux jeux d'argent. *Archives de politique criminelle*, (1), 79-91.
- Franck, E., Verbeek, E., & Nüesch, S. (2009). Inter-market arbitrage in sports betting. NCER Working.
- Franck, E., Verbeek, E., & Nüesch, S. (2010). Prediction accuracy of different market structures—bookmakers versus a betting exchange. *International Journal of Forecasting*, 26(3), 448-459.
- Gil, R. G. R., & Levitt, S. D. (2012). Testing the efficiency of markets in the 2002 World Cup. *The Journal of Prediction Markets*, 1(3), 255-270.
- Gillet, P. (1999). L'efficience des marchés financiers. *Economica*.
- Goddard, J., & Asimakopulos, I. (2004). Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting*, 23(1), 51-66.
- Goddard, J., & Asimakopulos, I. (2004). Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting*, 23(1), 51-66.
- Golec, J., & Tamarkin, M. (1991). The degree of inefficiency in the football betting market. *Journal of Financial Economics*, 30(2), 311-323.
- Gray, P. K., & Gray, S. F. (1997). Testing market efficiency: Evidence from the NFL sports betting market. *The Journal of Finance*, 52(4), 1725-1737.
- Griffith, R. M. (1949). Odds adjustments by American horse-race bettors. *The American Journal of Psychology*, 62(2), 290-294.
- Hausch, D. B., & Ziemba, W. T. (1995). Efficiency of sports and lottery betting markets. *handbooks in Operations research and Management Science*, 9, 545-580.
- Humphreys, B. R., & Soebbing, B. (2012). Sports betting, sports bettors and sports gambling policy. In *Sport und Sportgroßveranstaltungen in Europa-zwischen Zentralstaat und Regionen* (pp. 15-37). Hamburg: Hamburg Univ. Press.

- Jensen, M. C. (1978). Some anomalous evidence regarding market efficiency. *Journal of financial economics*, 6(2/3), 95-101.
- MacKenzie, D., & Millo, Y. (2003). Constructing a market, performing theory: The historical sociology of a financial derivatives exchange. *American journal of sociology*, 109(1), 107-145.
- Malios, W. S., (2000), *The analysis of sports forecasting: Modeling parallels between sports gambling and financial markets*, (Kluwer Academic Publishers, Boston)
- Malkiel, B. G. (1989). Efficient market hypothesis. In *Finance* (pp. 127-134). Palgrave Macmillan, London.
- Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2), 383-417.
- Mandelbrot, B. B. (1971). When can price be arbitrated efficiently? A limit to the validity of the random walk and martingale models. *The Review of Economics and Statistics*, 225-236.
- Mignon, V. (2008). Les ambiguïtés de la théorie de l'efficience informationnelle des marchés financiers. *Regards croisés sur l'économie*, (1), 104-117.
- Moroney, M. J. (1956). Facts from figures.
- Muniesa, F. (2000). Performing prices: The case of price discovery automation in the financial markets.
- Muniesa, F. (2003). *Des marchés comme algorithmes: sociologie de la cotation électronique à la Bourse de Paris* (Doctoral dissertation, École Nationale Supérieure des Mines de Paris).
- Nyberg, H. (2014). A multinomial logit-based statistical test of association football betting market efficiency.
- Ottaviani, M., & Sørensen, P. N. (2005). Parimutuel versus fixed-odds markets. *Unpublished Paper*.
- Ottaviani, M., & Sørensen, P. N. (2008). The favorite-longshot bias: An overview of the main explanations. In *Handbook of Sports and Lottery markets* (pp. 83-101). Elsevier.
- Pavée, F. (2010). *Paris sportifs en ligne: Mode d'emploi*. France: RMC Sports.Toutlemonde 2011
- Peel, D.A., & Pope, P. F. (1989). Information, prices and efficiency in a fixed-odds betting market. *Economica*, 323-341.
- Plott, C. R., Wit, J., & Yang, W. C. (2003). Parimutuel betting markets as information aggregation devices: experimental results. *Economic Theory*, 22(2), 311-351.
- Pope, P. F., & Peel, D. A. (1989). Information, prices and efficiency in a fixed-odds betting market. *Economica*, 323-341.

- Quandt, R. E. (1986). Betting and equilibrium. *The Quarterly Journal of Economics*, 101(1), 201-207.
- Rhode, P. W., & Strumpf, K. S. (2004). Historical presidential betting markets. *Journal of Economic Perspectives*, 18(2), 127-141.
- Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of finance*, 39(4), 1127-1139.
- Ruhm, D. L. (2003). Distribution-based pricing formulas are not arbitrage-free. In *Proceedings of the Casualty Actuarial Society* (Vol. 90).
- Samuelson, Paul (1965). "Proof That Properly Anticipated Prices Fluctuate Randomly". *Industrial Management Review*. 6: 41–49
- Sauer, R. D. (1998). The economics of wagering markets. *Journal of economic Literature*, 36(4), 2021-2064.
- Schnytzer, A., & Shilony, Y. (1995). Inside information in a betting market. *The Economic Journal*, 963-971
- Smith, M. A., Paton, D., & Williams, L. V. (2006). Market efficiency in person-to-person betting. *Economica*, 73(292), 673-689.
- Smith, M. A., Paton, D., & Williams, L. V. (2006). Market efficiency in person-to-person betting. *Economica*, 73(292), 673-689.
- Smith, M. A., Paton, D., & Williams, L. V. (2009). Do bookmakers possess superior skills to bettors in predicting outcomes?. *Journal of Economic Behavior & Organization*, 71(2), 539-549.
- Tauchen, G., Zhang, H., & Liu, M. (1996). Volume, volatility, and leverage: A dynamic analysis. *Journal of Econometrics*, 74(1), 177-208.
- Thaler, R. H., & Ziemba, W. T. (1988). Anomalies: Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic perspectives*, 2(2), 161-174.
- Trespeuch, M. (2011). Du monopole à la concurrence sur Internet. *Revue française de socio-économie*, (2), 39-58.
- Van Rompuy, B. (2015). The Odds of Match Fixing-Facts & Figures on the Integrity Risk of Certain Sports Bets.
- Vecer, J., Ichiba, T., & Laudanovic, M. (2006). Parallels between betting contracts and credit derivatives: Lessons learned from fifa world cup 2006 betting markets. *Science*, 1-15.
- Vergin, R. C., & Scriabin, M. (1978). Winning strategies for wagering on National Football League games. *Management Science*, 24(8), 809-818.



Vergin, R. C., & Sosik, J. J. (1999). No place like home: an examination of the home field advantage in gambling strategies in NFL football. *Journal of Economics and Business*, 51(1), 21-31.

Vlastakis, N., Dotsis, G., & Markellos, R. N. (2009). How efficient is the European football betting market? Evidence from arbitrage and trading strategies. *Journal of Forecasting*, 28(5), 426-444.




Wang, Y. (2003). Nonparametric tests for randomness. *ECE*, 461, 1-11.

Williams, L. V. (1999). Information efficiency in betting markets: A survey. *Bulletin of Economic Research*, 51(1), 1-39.

Wolfers, J., & Zitzewitz, E. (2004). Prediction markets. *Journal of economic perspectives*, 18(2), 107-126.

## Annexes

A-1

Today 6 February				Winner FT		
	Barcelona Real Madrid	Today 21:00	 	Barcelona <b>1.74</b>	Draw <b>4.05</b>	Real Madrid <b>4.40</b>
						+175 >